



King's Research Portal

DOI:

[10.1016/j.jaci.2020.09.042](https://doi.org/10.1016/j.jaci.2020.09.042)

[Link to publication record in King's Research Portal](#)

Citation for published version (APA):

Flohr, C. (2021). Gut microbiota development during infancy: impact of introducing allergenic foods. *Journal of Allergy and Clinical Immunology*, 147(2), 613-621.e9. [e9]. <https://doi.org/10.1016/j.jaci.2020.09.042>

Citing this paper

Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

General rights

Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Research Portal

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

The Journal of Allergy and Clinical Immunology

Gut microbiota development during infancy: impact of introducing allergenic foods

--Manuscript Draft--

Manuscript Number:	JACI-D-20-00041R3
Article Type:	Original Article
Section/Category:	Mechanisms of allergy/immunology
Keywords:	atopic dermatitis, bacteria, diet, environment, food, microbiome, colonization, tolerance
Corresponding Author:	Carsten Flohr, MD PhD MSc MA MPhil FRCPCH FRCP St Thomas' Hospital and King's College London London, UNITED KINGDOM
First Author:	Tom Marrs, PhD
Order of Authors:	Tom Marrs, PhD Jay-Hyun Jo, PhD Michael R. Perkin, PhD D.W. Rivett, PhD Adam A. Witney, PhD K.D. Bruce, PhD Kirsty Logan, PhD Joanna Craven, MPH Suzana Radulovic, MD Versteeg SA, BSc Ronald van Ree, PhD WH Irwin McLean, FRS David P. Strachan, MD Gideon Lack, MD Heidi H. Kong, MD, MHSc Carsten Flohr, MD PhD MSc MA MPhil FRCPCH FRCP
Manuscript Region of Origin:	UNITED KINGDOM
Abstract:	<p>Background: The gut microbiota potentially plays an important role in the immunological education of the host during early infancy.</p> <p>Objective: We sought to determine how the infant gut microbiota evolve during infancy, particularly in relation to hygiene-related environmental factors, allergic and atopic disorders, and a randomized introduction of allergenic solids.</p> <p>Methods: 1,303 exclusively breastfed infants were enrolled in a dietary randomized controlled trial (EAT study) from three months of age. In this nested longitudinal study, fecal samples were collected at baseline with additional sampling of selected cases and controls at six and twelve months to study the evolution of their gut microbiota, using 16S ribosomal RNA gene targeted amplicon sequencing.</p> <p>Results: In the 288 baseline samples from exclusively breastfed infant at three months of age, the gut microbiota were highly heterogenous, forming three distinct clusters: Bifidobacterium -rich, Bacteroides -rich and Escherichia / Shigella -rich. Mode of delivery was the major discriminating factor. Increased Clostridium sensu stricto relative abundances at three months of age were associated with presence of atopic dermatitis (AD) on examination at three and twelve months of age. From the selected cases and controls with longitudinal samples (n=70), transition to Bacteroides -rich communities and influx of adult-specific microbes were observed during the first year of life. The introduction of allergenic solids promoted a significant increase in Shannon</p>

	<p>diversity and increased abundances of specific microbes, such as genera belonging to Prevotellaceae and Proteobacteria (e.g. Escherichia/Shigella), as compared to infants recommended to exclusively breastfeed.</p> <p>Conclusions: Specific gut microbiota of three-month-old breastfed infants were associated with cesarean birth, while increased Clostridium sensu stricto was associated with AD. The randomized introduction of allergenic solids from three months of age alongside breastfeeding was associated with differential dynamics of maturation of the gut microbial communities.</p>
--	---

Dear Editors,

We were delighted to hear that our manuscript has been provisionally accepted, subject to responding to a further comment from Reviewer 1.

Please see below.

Sincerely yours, Carsten Flohr

REVIEWER #1

Original comment/response:

"Please show the early vs standard introduction groups Shannon-diversity at 3 months, 6 months and 12 months (not the increase). You are describing a randomized trial, so we assume the 3 months baseline samples are similar between the groups. If this is not the case this should still be reported along with the increase analyses.

Response: Thank you for the suggestion. To allow evaluation of Shannon diversities between the groups, we have re-incorporated the initial submission figure comparing Shannon diversities at 3m, 6m, and 12m between the standard and early introduction groups, which is now Figure E7A."

Perhaps I wasn't clear in my first request. I would like to see comparisons of intervention groups at 3, 6 and 12 months. 3 individual boxplots (one for 3 month, one for 6 months, and one for 1 yr) showing Shannon index on y-axis and intervention group at x-axis and Wilcoxon tests for differences at each timepoint separately. This is important as it seems like your 3 month baseline Shannon may differ between intervention groups.

Response:

Thank you for the clarification. To allow evaluation of Shannon diversities between the intervention groups, we have changed the layout of Figure E7A. As Reviewer 1 noted, the

Shannon diversity in the Early Introduction group appears lower than the Standard Introduction group at three months. However, it does not meet statistical significance (unpaired Wilcoxon test). Additionally, there is no statistically significant difference between Early and Standard Introduction groups at six months. This, together with other analyses, indicate that while there is a significant increase in Shannon diversity from three to six months with early introduction of allergenic foods (paired Wilcoxon test), the impact of this change is not notable enough to differentiate the Early Introduction group from the Standard Introduction group with community alpha diversity at six months of age. Considering the strengths of this study (tightly controlled diet and longitudinal follow up of each individual) the longitudinal change between three to six months of each individual in the Early Introduction group remains of interest.

We have also amended the text in the manuscript accordingly. See page 21, line 10:

“Of note, six-month gut microbiota Shannon diversities were not statistically significantly different between the early and standard introduction groups, indicating that the increased microbial community diversity after early introduction of allergenic foods was comparable to exclusively breast-fed infants (Fig E7, A). However, the combined findings from this study demonstrate that early introduction of solid foods elicits considerable and consistent changes in gut microbial composition and diversity from three to six months of infancy.”

ORIGINAL ARTICLE

Title: Gut microbiota development during infancy: impact of introducing allergenic foods

Tom Marrs, PhD^{a,b,§}, Jay-Hyun Jo, PhD^{c,§}, Michael R. Perkin, PhD^d, D.W. Rivett, PhD^e, Adam A. Witney, PhD^f, K.D. Bruce, PhD^g, Kirsty Logan, PhD^a, Joanna Craven, MPH^a, Suzana Radulovic, MD^{a,b}, Versteeg SA, BSc^h, Ronald van Ree, PhD^{h,i}, WH Irwin McLean, FRS^j, David P. Strachan, MD^d, Gideon Lack, MD^a, Heidi H. Kong, MD, MHSc^{c,*} & Carsten Flohr, PhD^{k,*}

§ - Joint first authors

* - Joint last authors

^a Paediatric Allergy Research Group, Department of Women and Children's Health, School of Life Course Sciences, Faculty of Life Sciences and Medicine, King's College London, WC2R 2LS

^b Children's Allergies Department, Guy's and St Thomas' NHS Foundation Trust, St Thomas' Hospital, Westminster Bridge Road, Lambeth, SE1 7EH

^c Dermatology Branch, National Institute of Arthritis and Musculoskeletal and Skin Diseases, NIH, Bethesda, Maryland, USA

^d Population Health Research Institute, St George's, University of London, Cranmer Terrace, London, SW17 0RE

^e Department of Natural Sciences, Manchester Metropolitan University, Manchester, UK

^f Institute for Infection and Immunity, St George's, University of London, Cranmer Terrace, London, SW17 0RE

^g Molecular Microbiology Research Laboratory, Pharmaceutical Science Division, King's College London, London, UK

^h Experimental Immunology, Amsterdam University Medical Center, Amsterdam, The Netherlands

ⁱ Department of Otorhinolaryngology, Amsterdam University Medical Center, Amsterdam, The Netherlands

^j Division of Biological Chemistry and Drug Discovery, School of Life Sciences, University of Dundee, Dundee DD1 5EH, UK

^k Unit for Population-Based Dermatology Research, St John's Institute of Dermatology, School of Basic and Medical Biosciences, King's College London, UK

Corresponding author:

Professor Carsten Flohr, Unit for Population-Based Dermatology Research, St John's Institute of Dermatology, School of Basic and Medical Biosciences, King's College London and Guy's and St Thomas' NHS Foundation Trust, London, UK.

Telephone number: +44-(0)20 7188 7188, extension 51162

Fax number: +44-(0)20 7188 6379

Email: carsten.flohr@kcl.ac.uk

Funding: The main components of the EAT Study were jointly funded by the UK Food Standards Agency (FSA, grant code T07051) and the Medical Research Council (MRC, grant MC_G1001205). The skin-related aspects of the study were supported by a Clinician Scientist

Award from the UK National Institute for Health Research (NIHR) held by CF (NIHRCS/01/2008/009). The British Skin Foundation funded the fecal sample sequencing work (BSF large grant 5047i, Principle Investigator CF). TM, CF and GL are supported by the NIHR Biomedical Research Centre at Guy's & St Thomas' NHS Foundation Trust and King's College London. The McLean lab is funded by Wellcome Trust Programme (092530/Z/10/Z; to WHIM) and Bioresources grants (090066/B/09/Z; to WHIM). The Centre for Dermatology and Genetic Medicine, University of Dundee, is supported by a Wellcome Trust Strategic Award (098439/Z/12/Z; to WHIM). This work was supported by the Intramural Research Program of the US National Institutes of Health, National Institute of Arthritis and Musculoskeletal and Skin Diseases (J-HJ, HHK). J-HJ was also supported by a grant of the Korean Health Technology R&D Project, Ministry of Health and Welfare, Republic of Korea (HI15C1095). The views expressed in this publication are those of the authors and not necessarily those of the Food Standards Agency, the Medical Research Council, the NHS, the UK National Institute for Health Research, the UK Department of Health, the US National Institutes of Health or the Wellcome Trust.

Conflict of interest: GL has received research funding from ALK Abello, the Immune Tolerance Network, US National Peanut Board, Food Standards Agency, Medical Research Council, Food Allergy Initiative, Action Medical Research and Guys and St Thomas' Charity. GL has received sponsorship from Novartis, Sodilac, Spanish Society of Allergy and Clinical Immunology, Danone Nutricia and Nestle. GL has been a scientific advisor for DBV Technologies and gives voluntary advice to the Anaphylaxis Campaign and US National Peanut Board. All other authors have no conflicts of interest.

Word count: 3201 words

Number of figures: 4 in main article, 5 in supplementary appendix

Abstract

Background: The gut microbiota potentially plays an important role in the immunological education of the host during early infancy.

Objective: We sought to determine how the infant gut microbiota evolve during infancy, particularly in relation to hygiene-related environmental factors, allergic and atopic disorders, and a randomized introduction of allergenic solids.

Methods: 1,303 exclusively breastfed infants were enrolled in a dietary randomized controlled trial (EAT study) from three months of age. In this nested longitudinal study, fecal samples were collected at baseline with additional sampling of selected cases and controls at six and twelve months to study the evolution of their gut microbiota, using *16S ribosomal RNA* gene targeted amplicon sequencing.

Results: In the 288 baseline samples from exclusively breastfed infant at three months of age, the gut microbiota were highly heterogenous, forming three distinct clusters: *Bifidobacterium*-rich, *Bacteroides*-rich and *Escherichia/Shigella*-rich. Mode of delivery was the major discriminating factor. Increased *Clostridium sensu stricto* relative abundances at three months of age were associated with presence of atopic dermatitis (AD) on examination at three and twelve months of age. From the selected cases and controls with longitudinal samples (n=70), transition to *Bacteroides*-rich communities and influx of adult-specific microbes were observed during the first year of life. The introduction of allergenic solids promoted a significant increase in Shannon diversity and increased abundances of specific microbes, such as genera belonging to *Prevotellaceae* and Proteobacteria (e.g. *Escherichia/Shigella*), as compared to infants recommended to exclusively breastfeed.

1 **Conclusions:** Specific gut microbiota of three-month-old breastfed infants were associated with
2 cesarean birth, while increased *Clostridium sensu stricto* was associated with AD. The
3 randomized introduction of allergenic solids from three months of age alongside breastfeeding
4 was associated with differential dynamics of maturation of the gut microbial communities.
5

Key messages:

- Cesarean delivery is associated with reduced gut microbiota diversity at three months.
- The infant gut microbiomes mature towards *Bacteroides*-rich communities that begin to transition during the latter half of the first year of life.
- Early introduction of solids into babies' diets accelerates microbiota diversity and promotes different trajectory of maturation.
- *Clostridium sensu strictu* abundance at three months was associated with AD and AD severity.

Capsule summary:

Among exclusively breastfed infants, the introduction of allergenic solids alongside breastfeeding was associated with a differential maturation of the gut microbiota.

Keywords: atopic dermatitis, bacteria, diet, environment, food, microbiome, colonization, tolerance

1 **Abbreviations:**

2	AD	Atopic dermatitis
3	CI	Confidence Interval
4	cOR	Crude Odds Ratio
5	EAT Study	Enquiring About Tolerance Study
6	OTU	Operational Taxonomic Unit
7	rRNA	Ribosomal ribonucleic acid
8	TEWL	Transepidermal water loss

1 Introduction

2 Patterns of gut microbiota acquisition during infancy, especially during the first year, are highly
3 dynamic and associated with the development of various conditions in later life, such as atopic
4 diseases (1-4). Mechanisms for how early colonization of differential microbes influences health
5 and diseases are under active research, and evidence has suggested that colonization of microbes
6 is indispensable for immune maturation and organ development. Major events – mode of
7 delivery, feeding patterns, and antibiotic usage – have been shown to affect the early microbiota
8 acquisition (5-11). Although many studies have been conducted to explore the effects of
9 environmental factors on gut microbiome development, most of them are retrospective and
10 observational, therefore, challenged by mixed confounders.

11 Mode of delivery has been identified as the prominent factor during the newborn period, as it
12 determines colonizing microorganisms that are shared from mother to newborn and may also
13 influence gut microbiota characteristics (5-7, 12). It has been commonly reported that cesarean
14 delivered infants harbour less *Bacteroides* and more hospital-associated pathobionts, such as
15 *Enterococcus* spp. and *Klebsiella* spp. An infant's diet may also influence its gut microbiota.
16 Several reports have highlighted gut microbiota differences between breastfed and formula-fed
17 infants (8, 13, 14). Two small uncontrolled studies have described an increase in infants' gut
18 microbiota complexity when dietary solids are introduced (15, 16), and a non-randomized study
19 of 98 maternal-infant dyads observed that cessation of breastfeeding was associated with
20 maturation into an adult-like microbiota (8).

21 Here, we investigated a nested cohort of infants undergoing randomized introduction of
22 allergenic solids as part of a randomized controlled trial to prevent food allergy. This study

- 1 provided an opportunity to evaluate the effects of early introduction of solid foods and other
- 2 factors including mode of delivery and environmental exposures on the shifts and maturation of
- 3 the gut microbiota and the risk of developing atopic disorders (17, 18).

1 **Methods**

2 EAT cohort

3 In the EAT study, 1303 exclusively breastfed, healthy infants from England and Wales, aged
4 between 12 and 17 weeks, were enrolled into a randomized controlled trial to examine
5 whether the regular consumption of allergenic solids reduced the prevalence of food allergy
6 when compared to continued exclusive breastfeeding up to six months of age (ISRCTN:
7 14254740) (17). Participants were randomly assigned to either the regular consumption of six
8 allergenic foods (boiled hen's egg, peanut, cow's milk (yoghurt), wheat, white fish and sesame)
9 beginning at three months of age twice weekly alongside continued breastfeeding (early
10 introduction group) or exclusive breastfeeding until six months (standard introduction group).
11 Beyond six months, food consumption was at parental discretion. Food specific IgE to each of
12 the six foods was measured at enrollment, one and three years of age in both groups using
13 ImmunoCap (Phadia) assays (cut off ≥ 0.35 kU/L to determine food sensitization). The primary
14 outcome of the original EAT Study was the prevalence of challenge-proven food allergy to one
15 or more of the six study foods between one and three years of age. Parents completed online
16 questionnaires eliciting exposure data, such as mode of delivery, sibship size, pet ownership,
17 and antibiotic usage.

18 All infants were examined for atopic dermatitis (AD) at their enrolment visit at three months of
19 age and twelve months, using the UK diagnostic criteria-based photographic protocol of the
20 International Study of Asthma and Allergies in Childhood (ISAAC) Phase Two (19). AD severity was
21 determined by the Scoring Atopic Dermatitis (SCORAD) index (20). Skin barrier function was
22 assessed by measuring trans-epidermal water loss (TEWL) using the Biox Aquaflux AF200 (Biox,

London, UK) closed condenser chamber device on unaffected skin of the left volar forearm. Venous blood samples were screened for the 6 commonest *FLG* mutations (TaqMan allelic discrimination assays for mutations R501X, 2282del4, R2447X, S3247X; ABI 7900 HT; Applied Biosystems, Foster City, Calif) or by sizing of fluorescent PCR products on an Applied Biosystems 3130 DNA sequencer (mutations 3673delC and 3702delG). These 6 mutations detect 99% of *FLG* mutation carriers in the UK population.

Ethical approval for the EAT Study was provided by the St Thomas' Hospital Research Ethics Committee (REC Reference 08/H0802/93) and informed consent was obtained from the parents of all children enrolled in the study.

Gut microbiota sample collection, DNA extraction and sequencing

Beginning the 1st of October 2011, consecutively enrolling families (n=359) were asked to provide stool samples (see Fig E1). Stool samples were collected from 288 of enrolled participants with follow up samples requested at six and twelve months of age. Stool samples were retrieved from diapers and transferred to the King's College London Molecular Microbiology Research Laboratory for storage at -80°C until DNA extraction. The cell lysis and DNA extraction protocol utilized both phenol-chloroform liquid-liquid phase separation and bead-beating, followed by -80°C storage (see Supplementary Appendix). To evaluate the assembly of gut microbial communities during infancy, a total of 428 samples were sequenced (of the total 288 subjects providing baseline samples, 218 participants provided baseline samples only, and a subset of 70 individuals provided samples from multiple time points of three, six, and twelve months). The subset of cases and controls (42 in the standard-introduction group and 28 in the early-

introduction group) were selected based on intervention assignment and atopy (either challenge-proven food allergy or AD on examination) status. DNA extracts were amplified using primers targeting the hypervariable V4 (515F-806R) region of the 16S ribosomal RNA gene and performed on the MiSeq instrument (21).

Sequence analysis

The open-source software mothur (22) and DADA2 (23) were used to process the raw sequencing data (see online supplement). Reads were aligned and classified down to the genus level with ribosomal database project naïve Bayesian classifier (24) or species level with Greengenes database v13.8 (25). Operational taxonomic units (OTUs) were defined at 97% similarity using the optclust algorithm (26). We also undertook comparative analyses with 977 samples from the TwinsUK (adults from the same UK region) (27, 28), which were sequenced using the same primers (Accession No. ERP006339 and ERP006342; 9 samples were removed during preprocessing) and TEDDY (large, multi-centre infant cohort) (11) studies. TEDDY dataset samples were selected based on i) exclusive breastfeeding before solid food introduction and ii) longitudinal data (≥ 2 samples before 6 months of age) to permit comparable analyses.

Statistical analysis

The beta-diversity and linear discriminant statistical analyses were performed using R software, with libraries including vegan, ape, and bios2mds. Thetayc distance (theta) metric was used to measure distance between communities (beta diversity) (29), and principal coordinate analysis

1 (PCoA) was performed based on theta distance, unless otherwise indicated. Significance of
2 differences between groups was tested by various methods, including Wilcoxon rank-sum,
3 Kruskal-wallis, and the analysis of molecular variance (AMOVA) (30). Covariates of community
4 variation were evaluated by calculating correlation between indicated PCoA ordination and
5 metadata or genus abundance (envfit function in vegan R package; 10,000 permutations with
6 Bonferroni correction).

Results

Of enrolled families, 80.2% (288 of 359) contributed baseline samples, a subset of which (n=70) also contributed follow-up samples. Within the subset, cases were comprised of twelve participants with allergies to egg (one additionally allergic to peanut), one milk allergy, and one codfish allergy, and thirteen with AD (SCORAD over 15 at either three or twelve months).

Exclusively breastfed infants' gut microbial communities form 3 distinct clusters

To understand the landscape of the gut microbiome during breastfeeding, we analyzed baseline (3-month) samples (n=288, and a mean age of 105 days and interquartile range 100-112 days). The most abundant genera established in the microbial communities at three months were *Bifidobacterium* (median abundance (MA): 38.75%), *Bacteroides* (MA: 5.43%) and multiple genera belonging to Firmicutes and Proteobacteria (Fig 1, A and B). The gut microbial communities of exclusively breastfed three-month-old infants was diverse, in examining the PCoA with theta distance (29) and Bray-Curtis distance based on genus-level classification (Fig 1, A and see Fig E2, A). k-means clustering (see Fig E2, A and B) revealed that the microbial communities of exclusively breastfed infants clustered into 3 groups, in which cluster with *Bifidobacterium*-rich cluster #1, *Bacteroides*-rich cluster #2 and cluster #3 predominated by *Escherichia/Shigella* and genera belonging to Firmicutes phylum (Fig 1, A and B). Although minor genera, such as *Streptococcus*, *Dorea* and *Rothia*, also contributed to the community variation, the three major genera were sufficient to reproduce clustering, underscoring their importance in community variation (Fig 1, C). Community diversity was lower in cluster #1 than in clusters #2 and #3 (alpha diversity; Shannon index, $p < 0.005$; Wilcoxon-rank sum test) due to the predominance of *Bifidobacterium* (Fig 1, D). Of note, the age of participants did not have a

significant effect on clustering (see Fig E2, C, $p > 0.05$). PCoA and k-means clustering based on theta distance calculated from species-level (see Fig E2, D) and *de novo* OTU (97% similarity) abundance (see Fig E2, E) and on Bray-Curtis distance (see Fig E2, A) were highly comparable to the genus-based clustering (Fig 1, B).

Infants' gut microbial communities mature in first year

Data from six- and twelve-month samples ($n=70$ samples per each time point; see methods) were superimposed onto the existing three-month PCoA coordinates (see Fig E3, A). Microbiota from six-month-old infants were similarly dispersed as three-month communities, whereas twelve-month samples were exclusively overlaid on cluster #2, indicating that gut microbial communities mature towards *Bacteroides*-rich communities at 1 year of age, in keeping with the previous literature (7, 10). Interpersonal similarity increased during aging (see Fig E3, B) due to the convergence to *Bacteroides*-rich communities. We then investigated whether the infant gut microbiomes matured towards an adult-like pattern. Since gut microbiomes can vary in adult populations from distinct geographic regions (31), we compared our dataset to a representative dataset of fecal microbiota from adults in a geographic region analogous to our infant cohort, revealing that infants begin to transition towards adult gut microbiomes during six to twelve months of age (see Fig E4). Collectively, these results demonstrated the extensive transition and maturation of gut microbiota during the first year of life.

Early environmental exposures are associated with the infant gut microbiota

Among the environmental exposures that were evaluated, mode of delivery was the most significant factor contributing to microbiome community variation ($p < 0.01$) at 3 months of age (Fig 2, A), with underrepresentation of *Bacteroides* in individuals born by cesarean section (Fig 2, B and C), as shown in previous reports (5, 32, 33). Since the mode of delivery profoundly contributed to microbiome variation, PCoA and clustering were re-analyzed after stratification by delivery methods. Gut microbial communities of vaginally delivered infants also formed 3 clusters, with the same three key genera – *Bifidobacterium*, *Bacteriodes*, and *Escherichia/Shigella* – as the major contributing taxa (see Fig E5, A), indicating contribution of other exposures in defining microbiome communities. Therefore, pair-wise associations between major bacterial genera (mean abundance $> 1\%$) and environmental factors were investigated (see Fig E5, B). Results highlighted additional bacterial genera associated with delivery mode (e.g. *Bifidobacterium*, *Clostridium sensu stricto*, Enterobacteriaceae, and Lachnospiraceae were enriched in infants with cesarean-delivered infants) underscoring the importance of delivery mode in shaping infant gut microbiomes.

Clinical outcomes are linked with specific infant gut microbiota

We next examined potential links between clinical outcomes and the 3-month gut microbiome clusters; however, no statistically significant associations were identified. Therefore, associations between major bacterial genera (mean abundance $> 1\%$) and clinical manifestations were also investigated (Fig, E6, A). Interestingly, SCORAD at 12 months of age was significantly associated with the abundance of *Clostridium sensu stricto* and *Haemophilus* at enrolment. Equally, raised TEWL at 3 months showed a positive association with *Haemophilus*. Raised TEWL at 12 months

was also linked to an increased abundance in *Veillonella* (Fig E6, A). Linear discriminant analysis confirmed higher relative abundances of *Clostridium sensu stricto* at three months of age were observed in infants who had AD at three months and twelve months of age (see Fig E6, B and C). Family history of allergic disease and food allergy/sensitization did not significantly differ according to the microbiome variation or clustering.

The introduction of allergenic foods alters infants' gut microbiota

The EAT Study randomized exclusively breastfed infants to dietary introduction of six allergenic foods ('early introduction', n=28) versus continued exclusive breastfeeding ('standard introduction', n=42) until 6 months of age (17). Therefore, the maturation in the gut microbiota was investigated relative to the timing of introduction of allergenic solids. Participants in the intervention group had a significantly increased Shannon microbial diversity index at six months as compared to their enrolment samples (Fig 3, A, and Fig E7, A). PCoA revealed the gut microbial communities of the early and standard introduction groups were largely indistinguishable (Fig 3, B, P non-significant; AMOVA) at three and twelve months. In contrast, these microbial communities significantly differed at six months of age (Fig 3, B, middle panel. $p < 0.05$; AMOVA), demonstrating the differential maturation trajectories of early versus standard introduction groups at six months of age. Interestingly, early intervention group infants predominantly transitioned along the second principal coordinate (y-axis) from three to six months of age (Fig 3, C). Abundances of *Prevotella* and *Escherichia/Shigella* were positively correlated with the second principal coordinate ($p < 0.05$, and $r^2 = 0.52$ and 0.44 , respectively). Similarly, linear discriminant analysis revealed *Paraprevotella* and various genera belonging to Proteobacteria and Firmicutes

1 phyla were enriched in the early intervention group (see Fig E7, B and D) at six months, but not
2 evident at 12 months of age (see Fig E7, C and E). Infants in both groups moved along with the
3 first principal coordinate (x-axis) and converged towards *Bacteroides*-rich communities during six
4 to twelve months (Fig 3, D). To validate that the introduction of solids altered the gut microbiota
5 in exclusively breastfed infants, we analyzed gut microbiome changes in exclusively breastfed
6 infants after similar dietary changes within the TEDDY cohort (see Methods for detailed criteria)
7 (11) and superimposed their data onto the same PCoA dimensions of (see E8, A). TEDDY study
8 infants with early introduction of solids (n=48, no formula and started solid foods before 150
9 days of life) also transitioned along with the second principal coordinate from 3 months to 6
10 months, whereas TEDDY study infants mimicking the standard introduction group (n=45, no
11 formula and started solid foods after 150 days of life) did not follow the same gut microbiome
12 transition (Fig 3, E).

1 Discussion

2 There remains deep interest in understanding what contribution external factors may contribute
3 to the development of atopic and allergic disorders. This cohort provided a unique opportunity
4 to examine the relationship between the gut microbiome and many factors including dietary
5 intervention, clinical outcomes, and hygiene-related factors. Gut microbiota are widely
6 heterogenous amongst exclusively breastfed three-month-old infants and have lesser abundance
7 of *Bacteroides* species if born by cesarean section. After three months of age, the gut microbiota
8 underwent major transformation in diversity, complexity, and predominance of 'adult-like'
9 bacteria. Babies' gut microbiota diversified when allergenic foods were introduced and matured
10 towards *Bacteroides*-rich communities at twelve months of age, with significant changes
11 observed at a younger age in infants with earlier introduction of allergenic solids beginning at
12 three months of age. While there were significant microbiota-related associations with AD and
13 AD severity at three and 12 months, none of the observed gut microbiota changes were
14 associated with the development of food allergy or sensitization.

15

16 *Study strengths and limitations*

17 Strengths of the EAT study cohort were strictly controlled eligibility of exclusively breastfed
18 infants, population-representative sampling from the general population across England and
19 Wales, and a randomized intervention with allergenic solid introduction between three- and six-
20 months age, rather than an observational study design with no control of food intake (17). The
21 nested observational study within the EAT cohort allowed comparisons of the effects of the

1 dietary intervention on the gut microbiota over the first year of life. We validated these findings
2 in exclusively breastfed infants with comparable introductions of solid foods in the TEDDY cohort,
3 recapitulating these observations in other populations with westernized diets. Detailed
4 phenotyping of this cohort enabled assessment of alterations in infant gut microbiota in
5 association with exposures, such as mode of delivery, antibiotic prescribing, communal child
6 care, sibship size, rural vs urban residence, and pet ownership. Validated scores were used to
7 examine participants for AD, and diagnosis of food allergy was challenge-based, rather than
8 relying on parental report. A limitation of our study is that we were unable to evaluate the gut
9 microbiota before three months of life, which would have allowed an even earlier
10 characterization of microbiota development.

11 *Neonatal environmental exposures may influence gut microbiota at three months of age*

13 Delivery by cesarean had the most profound effects on the shaping of the gut microbiota during
14 early infancy with reduction in *Bacteroides*, similar to other cohorts (5, 11, 32, 33). Colonization
15 with *Bacteroides* in the neonatal period may have considerable health benefits, especially
16 immune maturation (34). *Bacteroides fragilis* has Polysaccharide A molecular moieties which
17 stimulate the development of T-regulatory cells in the lamina propria (35) and samples from
18 human fecal microbiota have found that at least 16 *Bacteroides* taxa contribute to this
19 phenomenon *in vivo* (36). Therefore, the abundance of *Bacteroides* spp. may relate to the nature
20 of early life immunological priming and warrants further investigation. In addition, if these
21 mechanistic studies are fruitful, intervention trials may consider targeting infants born by
22 cesarean to try and alter *Bacteroides* relative abundances and/or *Bacteroides* diversity (7).

Although we did not find other exposures contributing to the heterogeneity of microbiome at three months of infancy, there may be mixed effects of multiple factors and/or hidden factors which were not included in our metadata; therefore, an optimized model for discovering associated factors might be necessary. With the current dataset, we trained a random forest classifier with metadata to predict 3 clusters at three months. While the classifier did not perform well (out-of-box error rate > 40%, data not shown), it calculated maternal age at birth as a second important factor for classifying clusters, suggesting the importance of maternal/household factors in microbiome development at infancy. Similarly, studies have reported the long-term colonization of vertically transmitted microbes (12).

Diet contributes to evolution of the infant gut microbiota

Infants' diet also had significant effects on the shaping of the gut microbiota during early infancy. Several reports have demonstrated differential microbiome colonization in breastfed versus formula-fed infants, with a higher prevalence of *Bifidobacterium* spp. (*B. longum* and *B. breve*) and *Lactobacillus* in breastfed infants (8, 9, 11, 14). In addition, the introduction of solid foods has been proposed as a key contributor to the maturation of the gut microbiota. For example, a Scandinavian observational study concluded that the gut microbiota matured after the cessation of breastfeeding (8).

Since studies randomizing dietary interventions in infants are very unusual, this study offered a unique opportunity to study the effects of introducing allergenic solids on the gut microbiome. In the EAT study, early peanut and egg introduction, if consumed in sufficient quantity, was

shown to protect against the development of peanut and egg allergies between one and three years of age (17, 18). We have demonstrated that the early introduction of allergenic foods alongside ongoing breastfeeding between three and six months of age led to an increase in overall gut microbiota Shannon diversity, in particular promoting an influx of various microbes including *Prevotellaceae* and *Escherichia/Shigella*. Interestingly, presence of *Prevotella* has been shown to be associated with high-fiber diet including remote villages with less frequent chronic inflammatory disorders (37-40). Additionally, preceding studies have noted that the cessation of breastfeeding may promote further diversification of infants' gut microbiota(8), whereas our results showed that the introduction of a variety of allergenic foods in the setting of exclusive breastfeeding increased diversity of the gut microbiota. For a note, there was no difference in six-month gut microbiota community Shannon diversities between the early and standard introduction groups at six months, indicating that the increased microbial community diversity of individuals in the after early introduction of allergenic foods was comparable to exclusively breast-fed infants group is not substantial enough to be reflected at a population level (Fig E7, A). However, with other analyses in current study, our results the combined findings from this study demonstrate that early introduction of solid foods elicits considerable and consistent changes in gut microbial composition and diversity during from three to six months of infancy. Therefore, controlled dietary interventions in infancy may promote selective colonization of desired microbes and potentially exerts health and developmental benefit to the host. However, we found no evidence that the observed *changes* in the gut microbiota driven by the early introduction of allergenic foods impact on the development of AD.

1 In conclusion, we confirmed that cesarean delivery is associated with reduced gut microbiota
2 diversity at three months. The early introduction of solids into infants' diets accelerates
3 maturation of microbiota diversity as well as increases the relative abundances of *Prevotella* and
4 *Escherichia/Shigella*. This manipulation of the gut microbiota in early life towards diversity and
5 maturity, in particular in those delivered by cesarean section or at risk of allergy and atopic
6 diseases, should be examined in future studies to examine potential wider health benefits.

1 **Acknowledgements**

2 We thank the parents and children for taking part in this trial; the members of the trial steering
3 committee (Graham Roberts [chair], David Strachan [vice chair], Mary Fewtrell, Christine
4 Edwards, David Reading, Ian Kimber, Anne Greenough, Andy Grieve, Mary Feeney, Kate
5 Grimshaw, Judy More, Debbie Palmer, Carina Venter, and Rebecca Knibb) for contributions to
6 the study design; Monica Basting and Gemma Deutsch for project management; Helen Fisher,
7 Una O'Dwyer-Leeson, Amy Nixon, Louise Coverdale, and Muhsinah Adam for nursing support;
8 Alicia Parr for dietetic support; George Du Toit and Susan Chan for assistance with medical
9 supervision; Jenna Heath and Kathryn Hersee for play-specialist support; and Joelle Buck, Sarah
10 Hardy, Elizabeth Kendall, and Shuhana Begum, of the Food Standards Agency, for their
11 commitment to the trial.

Data Access

All sequencing data were deposited and are available at the NCBI Sequence Read Archive (SRA) under BioProject number PRJNA597342.

References

1. Zeissig S, Blumberg RS. Life at the beginning: perturbation of the microbiota by antibiotics in early life and its role in health and disease. *Nat Immunol*. 2014;15(4):307-10.
2. Arrieta MC, Stiemsma LT, Dimitriu PA, Thorson L, Russell S, Yurist-Doutsch S, et al. Early infancy microbial and metabolic alterations affect risk of childhood asthma. *Sci Transl Med*. 2015;7(307):307ra152.
3. Fujimura KE, Sitarik AR, Havstad S, Lin DL, Levan S, Fadrosch D, et al. Neonatal gut microbiota associates with childhood multisensitized atopy and T cell differentiation. *Nat Med*. 2016;22(10):1187-91.
4. Stokholm J, Blaser MJ, Thorsen J, Rasmussen MA, Waage J, Vinding RK, et al. Maturation of the gut microbiome and risk of asthma in childhood. *Nat Commun*. 2018;9(1):141.
5. Dominguez-Bello MG, Costello EK, Contreras M, Magris M, Hidalgo G, Fierer N, et al. Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proc Natl Acad Sci U S A*. 2010;107(26):11971-5.
6. Stokholm J, Thorsen J, Chawes BL, Schjorring S, Krogfelt KA, Bonnelykke K, et al. Cesarean section changes neonatal gut colonization. *J Allergy Clin Immunol*. 2016;138(3):881-9 e2.
7. Bokulich NA, Chung J, Battaglia T, Henderson N, Jay M, Li H, et al. Antibiotics, birth mode, and diet shape microbiome maturation during early life. *Sci Transl Med*. 2016;8(343):343ra82.
8. Backhed F, Roswall J, Peng Y, Feng Q, Jia H, Kovatcheva-Datchary P, et al. Dynamics and Stabilization of the Human Gut Microbiome during the First Year of Life. *Cell Host Microbe*. 2015;17(6):852.
9. Baumann-Dudenhoeffer AM, D'Souza AW, Tarr PI, Warner BB, Dantas G. Infant diet and maternal gestational weight gain predict early metabolic maturation of gut microbiomes. *Nat Med*. 2018.
10. Yassour M, Vatanen T, Siljander H, Hamalainen AM, Harkonen T, Ryhanen SJ, et al. Natural history of the infant gut microbiome and impact of antibiotic treatment on bacterial strain diversity and stability. *Sci Transl Med*. 2016;8(343):343ra81.
11. Stewart CJ, Ajami NJ, O'Brien JL, Hutchinson DS, Smith DP, Wong MC, et al. Temporal development of the gut microbiome in early childhood from the TEDDY study. *Nature*. 2018;562(7728):583-8.
12. Shao Y, Forster SC, Tsaliki E, Vervier K, Strang A, Simpson N, et al. Stunted microbiota and opportunistic pathogen colonization in caesarean-section birth. *Nature*. 2019;574(7776):117-21.
13. Penders J, Thijs C, Vink C, Stelma FF, Snijders B, Kummeling I, et al. Factors influencing the composition of the intestinal microbiota in early infancy. *Pediatrics*. 2006;118(2):511-21.
14. Bezirtzoglou E, Tsiotsias A, Welling GW. Microbiota profile in feces of breast- and formula-fed newborns by using fluorescence in situ hybridization (FISH). *Anaerobe*. 2011;17(6):478-82.
15. Palmer C, Bik EM, DiGiulio DB, Relman DA, Brown PO. Development of the human infant intestinal microbiota. *PLoS Biol*. 2007;5(7):e177.

16. Koenig JE, Spor A, Scalfone N, Fricker AD, Stombaugh J, Knight R, et al. Succession of microbial consortia in the developing infant gut microbiome. *Proc Natl Acad Sci U S A*. 2011;108 Suppl 1:4578-85.
17. Perkin MR, Logan K, Marrs T, Radulovic S, Craven J, Flohr C, et al. Enquiring About Tolerance (EAT) study: Feasibility of an early allergenic food introduction regimen. *J Allergy Clin Immunol*. 2016;137(5):1477-86 e8.
18. Perkin MR, Logan K, Tseng A, Raji B, Ayis S, Peacock J, et al. Randomized Trial of Introduction of Allergenic Foods in Breast-Fed Infants. *N Engl J Med*. 2016;374(18):1733-43.
19. Weiland SK, Bjorksten B, Brunekreef B, Cookson WO, von Mutius E, Strachan DP, et al. Phase II of the International Study of Asthma and Allergies in Childhood (ISAAC II): rationale and methods. *Eur Respir J*. 2004;24(3):406-12.
20. Kunz B, Oranje AP, Labreze L, Stalder JF, Ring J, Taieb A. Clinical validation and guidelines for the SCORAD index: consensus report of the European Task Force on Atopic Dermatitis. *Dermatology*. 1997;195(1):10-9.
21. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, et al. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc Natl Acad Sci U S A*. 2011;108 Suppl 1:4516-22.
22. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol*. 2009;75(23):7537-41.
23. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods*. 2016;13(7):581-3.
24. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol*. 2007;73(16):5261-7.
25. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol*. 2006;72(7):5069-72.
26. Westcott SL, Schloss PD. OptiClust, an Improved Method for Assigning Amplicon-Based Sequence Data to Operational Taxonomic Units. *mSphere*. 2017;2(2).
27. Goodrich JK, Waters JL, Poole AC, Sutter JL, Koren O, Blekhman R, et al. Human genetics shape the gut microbiome. *Cell*. 2014;159(4):789-99.
28. Jackson MA, Verdi S, Maxan ME, Shin CM, Zierer J, Bowyer RCE, et al. Gut microbiota associations with common diseases and prescription medications in a population-based cohort. *Nat Commun*. 2018;9(1):2655.
29. Yue JC, Clayton MK. A similarity measure based on species proportions. *Communications in Statistics-Theory and Methods*. 2005;34(11):2123-31.
30. Excoffier L, Smouse PE, Quattro JM. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*. 1992;131(2):479-91.
31. Clemente JC, Pehrsson EC, Blaser MJ, Sandhu K, Gao Z, Wang B, et al. The microbiome of uncontacted Amerindians. *Sci Adv*. 2015;1(3).

32. Rutayisire E, Huang K, Liu Y, Tao F. The mode of delivery affects the diversity and colonization pattern of the gut microbiota during the first year of infants' life: a systematic review. *BMC Gastroenterol.* 2016;16(1):86.
33. Jakobsson HE, Abrahamsson TR, Jenmalm MC, Harris K, Quince C, Jernberg C, et al. Decreased gut microbiota diversity, delayed Bacteroidetes colonisation and reduced Th1 responses in infants delivered by caesarean section. *Gut.* 2014;63(4):559-66.
34. Kelly D, Campbell JI, King TP, Grant G, Jansson EA, Coutts AG, et al. Commensal anaerobic gut bacteria attenuate inflammation by regulating nuclear-cytoplasmic shuttling of PPAR-gamma and RelA. *Nat Immunol.* 2004;5(1):104-12.
35. Round JL, Mazmanian SK. Inducible Foxp3⁺ regulatory T-cell development by a commensal bacterium of the intestinal microbiota. *Proc Natl Acad Sci U S A.* 2010;107(27):12204-9.
36. Atarashi K, Tanoue T, Shima T, Imaoka A, Kuwahara T, Momose Y, et al. Induction of colonic regulatory T cells by indigenous Clostridium species. *Science.* 2011;331(6015):337-41.
37. Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, et al. Human gut microbiome viewed across age and geography. *Nature.* 2012;486(7402):222-7.
38. De Filippo C, Cavalieri D, Di Paola M, Ramazzotti M, Poullet JB, Massart S, et al. Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. *Proc Natl Acad Sci U S A.* 2010;107(33):14691-6.
39. Schnorr SL, Candela M, Rampelli S, Centanni M, Consolandi C, Basaglia G, et al. Gut microbiome of the Hadza hunter-gatherers. *Nat Commun.* 2014;5:3654.
40. Marungruang N, Tovar J, Bjorck I, Hallenius FF. Improvement in cardiometabolic risk markers following a multifunctional diet is associated with gut microbial taxa in healthy overweight and obese subjects. *Eur J Nutr.* 2018;57(8):2927-36.

List of figure captions**FIG 1. Gut microbiota characteristics at three months of age**

A, Principal coordinate analysis (PCoA) of the gut microbiome at 3 months. Pair-wise distances (theta distance) among all samples were calculated and two major axes (PC1 and PC2) from the multi-dimensional distance space were calculated and depicted on a scatter plot. Colors indicate different clusters, according to k-means clustering. Arrows and letters indicate specific genera significantly correlated with PCoA ordination ($p < 0.05$, lengths of arrows are proportion to R^2 , calculated from EnvFit function from R). Corresponding genera for letters are showing on the right panel of B.

B, Averaged stacked bar chart of infants' microbiome within each cluster. Genera that are significantly correlated with PCoA ordination are depicted as colored bar (legend on right) and others are merged as 'Other bacteria'.

C, Triangle plot showing relative abundances of the three key genera (*Bifidobacterium*, *Bacteroides*, and *Escherichia/Shigella*) in the gut microbiota of baseline samples. The colors indicate clusters (same as Fig 1A).

D, Boxplot of bacterial community diversity (Shannon index) differences according to microbiome community clusters. Shannon diversity of Cluster #1 community is significantly lower than Cluster 2 and 3. (** $p < 0.01$, *** $p < 0.001$; Wilcoxon rank-sum test, after Kruskal-wallis test)

FIG 2. Evaluation of associations between gut microbiota characteristics and cesarean delivery

A, Association analysis of clinical, hygiene and life-style factors ('covariates') with microbiome variations. Among total of 48 covariates tested, 11 covariates with $r^2 > 0.01$ are plotted here.

Each bar indicates the amount of variance explained by each covariate, calculated by EnvFit function from R. (** $p < 0.01$; Bonferroni corrected) After correction for multiple comparison, mode of delivery remained as a significant factor that associated to microbiome variation.

B, PCoA of microbiomes at 3 months with delivery mode delineated. Color indicates cluster (same as Fig 1, A) and the shape indicates delivery mode. Arrows demonstrate direction of each covariate in the ordination space. Infants born by cesarean section tend to cluster on the left portion of scatter plot.

C, Boxplot showing the difference of relative abundances of *Bacteroides* by mode of delivery. *Bacteroides* are largely missing in infants born by cesarean section. (** $p < 0.001$; Wilcoxon rank-sum test)

FIG 3. Impact of allergenic solid introduction on infants' gut microbiota

A, Boxplot comparing Shannon diversity changes amongst participants' longitudinal samples according to randomized allocation to continued exclusive breastfeeding (standard introduction group) or the introduction of allergenic solids (early introduction group) (* $p < 0.05$; Wilcoxon rank-sum test).

B, PCoA scatter plot demonstrating longitudinal transition from 3 to 6 month of age. Grey arrows connect samples from the same individuals. Yellow and purple arrows on the sides indicate the average shift of the microbiota in each PCoA axis.

C, PCoA scatter plot demonstrating longitudinal transition from 6 to 12 month of age. Grey arrows connect samples from the same individual. Black arrows on the sides indicate average shift of the microbiota in each PCoA axis.

D, PCoA plot showing longitudinal transition of the gut microbiome from 3 to 12 months of age in the different dietary intervention groups. At 6 months, the microbiota of early introduction is significantly different from standard introduction ($p < 0.05$ by AMOVA).

E, Boxplot showing changes of microbiome from 3 to 6 months, in different dietary intervention groups and cohorts (EAT and TEDDY). Early introduction in both studies led to an increase along the second principal coordinate axis (* $p < 0.05$ and ** $p < 0.01$; paired Wilcoxon rank-sum test).

ORIGINAL ARTICLE

Title: Gut microbiota development during infancy: impact of introducing allergenic foods

Tom Marrs, PhD^{a,b,§}, Jay-Hyun Jo, PhD^{c,§}, Michael R. Perkin, PhD^d, D.W. Rivett, PhD^e, Adam A. Witney, PhD^f, K.D. Bruce, PhD^g, Kirsty Logan, PhD^a, Joanna Craven, MPH^a, Suzana Radulovic, MD^{a,b}, Versteeg SA, BSc^h, Ronald van Ree, PhD^{h,i}, WH Irwin McLean, FRS^j, David P. Strachan, MD^d, Gideon Lack, MD^a, Heidi H. Kong, MD, MHSc^{c,*} & Carsten Flohr, PhD^{k,*}

§ - Joint first authors

* - Joint last authors

^a Paediatric Allergy Research Group, Department of Women and Children's Health, School of Life Course Sciences, Faculty of Life Sciences and Medicine, King's College London, WC2R 2LS

^b Children's Allergies Department, Guy's and St Thomas' NHS Foundation Trust, St Thomas' Hospital, Westminster Bridge Road, Lambeth, SE1 7EH

^c Dermatology Branch, National Institute of Arthritis and Musculoskeletal and Skin Diseases, NIH, Bethesda, Maryland, USA

^d Population Health Research Institute, St George's, University of London, Cranmer Terrace, London, SW17 0RE

^e Department of Natural Sciences, Manchester Metropolitan University, Manchester, UK

^f Institute for Infection and Immunity, St George's, University of London, Cranmer Terrace, London, SW17 0RE

^g Molecular Microbiology Research Laboratory, Pharmaceutical Science Division, King's College London, London, UK

^h Experimental Immunology, Amsterdam University Medical Center, Amsterdam, The Netherlands

ⁱ Department of Otorhinolaryngology, Amsterdam University Medical Center, Amsterdam, The Netherlands

^j Division of Biological Chemistry and Drug Discovery, School of Life Sciences, University of Dundee, Dundee DD1 5EH, UK

^k Unit for Population-Based Dermatology Research, St John's Institute of Dermatology, School of Basic and Medical Biosciences, King's College London, UK

Corresponding author:

Professor Carsten Flohr, Unit for Population-Based Dermatology Research, St John's Institute of Dermatology, School of Basic and Medical Biosciences, King's College London and Guy's and St Thomas' NHS Foundation Trust, London, UK.

Telephone number: +44-(0)20 7188 7188, extension 51162

Fax number: +44-(0)20 7188 6379

Email: carsten.flohr@kcl.ac.uk

Funding: The main components of the EAT Study were jointly funded by the UK Food Standards Agency (FSA, grant code T07051) and the Medical Research Council (MRC, grant MC_G1001205). The skin-related aspects of the study were supported by a Clinician Scientist

Award from the UK National Institute for Health Research (NIHR) held by CF (NIHRCS/01/2008/009). The British Skin Foundation funded the fecal sample sequencing work (BSF large grant 5047i, Principle Investigator CF). TM, CF and GL are supported by the NIHR Biomedical Research Centre at Guy's & St Thomas' NHS Foundation Trust and King's College London. The McLean lab is funded by Wellcome Trust Programme (092530/Z/10/Z; to WHIM) and Bioresources grants (090066/B/09/Z; to WHIM). The Centre for Dermatology and Genetic Medicine, University of Dundee, is supported by a Wellcome Trust Strategic Award (098439/Z/12/Z; to WHIM). This work was supported by the Intramural Research Program of the US National Institutes of Health, National Institute of Arthritis and Musculoskeletal and Skin Diseases (J-HJ, HHK). J-HJ was also supported by a grant of the Korean Health Technology R&D Project, Ministry of Health and Welfare, Republic of Korea (HI15C1095). The views expressed in this publication are those of the authors and not necessarily those of the Food Standards Agency, the Medical Research Council, the NHS, the UK National Institute for Health Research, the UK Department of Health, the US National Institutes of Health or the Wellcome Trust.

Conflict of interest: GL has received research funding from ALK Abello, the Immune Tolerance Network, US National Peanut Board, Food Standards Agency, Medical Research Council, Food Allergy Initiative, Action Medical Research and Guys and St Thomas' Charity. GL has received sponsorship from Novartis, Sodilac, Spanish Society of Allergy and Clinical Immunology, Danone Nutricia and Nestle. GL has been a scientific advisor for DBV Technologies and gives voluntary advice to the Anaphylaxis Campaign and US National Peanut Board. All other authors have no conflicts of interest.

Word count: 3201 words

Number of figures: 4 in main article, 5 in supplementary appendix

Abstract

Background: The gut microbiota potentially plays an important role in the immunological education of the host during early infancy.

Objective: We sought to determine how the infant gut microbiota evolve during infancy, particularly in relation to hygiene-related environmental factors, allergic and atopic disorders, and a randomized introduction of allergenic solids.

Methods: 1,303 exclusively breastfed infants were enrolled in a dietary randomized controlled trial (EAT study) from three months of age. In this nested longitudinal study, fecal samples were collected at baseline with additional sampling of selected cases and controls at six and twelve months to study the evolution of their gut microbiota, using *16S ribosomal RNA* gene targeted amplicon sequencing.

Results: In the 288 baseline samples from exclusively breastfed infant at three months of age, the gut microbiota were highly heterogenous, forming three distinct clusters: *Bifidobacterium*-rich, *Bacteroides*-rich and *Escherichia/Shigella*-rich. Mode of delivery was the major discriminating factor. Increased *Clostridium sensu stricto* relative abundances at three months of age were associated with presence of atopic dermatitis (AD) on examination at three and twelve months of age. From the selected cases and controls with longitudinal samples (n=70), transition to *Bacteroides*-rich communities and influx of adult-specific microbes were observed during the first year of life. The introduction of allergenic solids promoted a significant increase in Shannon diversity and increased abundances of specific microbes, such as genera belonging to *Prevotellaceae* and Proteobacteria (e.g. *Escherichia/Shigella*), as compared to infants recommended to exclusively breastfeed.

1 **Conclusions:** Specific gut microbiota of three-month-old breastfed infants were associated with
2 cesarean birth, while increased *Clostridium sensu stricto* was associated with AD. The
3 randomized introduction of allergenic solids from three months of age alongside breastfeeding
4 was associated with differential dynamics of maturation of the gut microbial communities.
5

Key messages:

- Cesarean delivery is associated with reduced gut microbiota diversity at three months.
- The infant gut microbiomes mature towards *Bacteroides*-rich communities that begin to transition during the latter half of the first year of life.
- Early introduction of solids into babies' diets accelerates microbiota diversity and promotes different trajectory of maturation.
- *Clostridium sensu strictu* abundance at three months was associated with AD and AD severity.

Capsule summary:

Among exclusively breastfed infants, the introduction of allergenic solids alongside breastfeeding was associated with a differential maturation of the gut microbiota.

Keywords: atopic dermatitis, bacteria, diet, environment, food, microbiome, colonization, tolerance

1 **Abbreviations:**

2	AD	Atopic dermatitis
3	CI	Confidence Interval
4	cOR	Crude Odds Ratio
5	EAT Study	Enquiring About Tolerance Study
6	OTU	Operational Taxonomic Unit
7	rRNA	Ribosomal ribonucleic acid
8	TEWL	Transepidermal water loss

1 Introduction

2 Patterns of gut microbiota acquisition during infancy, especially during the first year, are highly
3 dynamic and associated with the development of various conditions in later life, such as atopic
4 diseases (1-4). Mechanisms for how early colonization of differential microbes influences health
5 and diseases are under active research, and evidence has suggested that colonization of microbes
6 is indispensable for immune maturation and organ development. Major events – mode of
7 delivery, feeding patterns, and antibiotic usage – have been shown to affect the early microbiota
8 acquisition (5-11). Although many studies have been conducted to explore the effects of
9 environmental factors on gut microbiome development, most of them are retrospective and
10 observational, therefore, challenged by mixed confounders.

11 Mode of delivery has been identified as the prominent factor during the newborn period, as it
12 determines colonizing microorganisms that are shared from mother to newborn and may also
13 influence gut microbiota characteristics (5-7, 12). It has been commonly reported that cesarean
14 delivered infants harbour less *Bacteroides* and more hospital-associated pathobionts, such as
15 *Enterococcus* spp. and *Klebsiella* spp. An infant's diet may also influence its gut microbiota.
16 Several reports have highlighted gut microbiota differences between breastfed and formula-fed
17 infants (8, 13, 14). Two small uncontrolled studies have described an increase in infants' gut
18 microbiota complexity when dietary solids are introduced (15, 16), and a non-randomized study
19 of 98 maternal-infant dyads observed that cessation of breastfeeding was associated with
20 maturation into an adult-like microbiota (8).

21 Here, we investigated a nested cohort of infants undergoing randomized introduction of
22 allergenic solids as part of a randomized controlled trial to prevent food allergy. This study

- 1 provided an opportunity to evaluate the effects of early introduction of solid foods and other
- 2 factors including mode of delivery and environmental exposures on the shifts and maturation of
- 3 the gut microbiota and the risk of developing atopic disorders (17, 18).

1 **Methods**

2 EAT cohort

3 In the EAT study, 1303 exclusively breastfed, healthy infants from England and Wales, aged
4 between 12 and 17 weeks, were enrolled into a randomized controlled trial to examine
5 whether the regular consumption of allergenic solids reduced the prevalence of food allergy
6 when compared to continued exclusive breastfeeding up to six months of age (ISRCTN:
7 14254740) (17). Participants were randomly assigned to either the regular consumption of six
8 allergenic foods (boiled hen's egg, peanut, cow's milk (yoghurt), wheat, white fish and sesame)
9 beginning at three months of age twice weekly alongside continued breastfeeding (early
10 introduction group) or exclusive breastfeeding until six months (standard introduction group).
11 Beyond six months, food consumption was at parental discretion. Food specific IgE to each of
12 the six foods was measured at enrollment, one and three years of age in both groups using
13 ImmunoCap (Phadia) assays (cut off ≥ 0.35 kU/L to determine food sensitization). The primary
14 outcome of the original EAT Study was the prevalence of challenge-proven food allergy to one
15 or more of the six study foods between one and three years of age. Parents completed online
16 questionnaires eliciting exposure data, such as mode of delivery, sibship size, pet ownership,
17 and antibiotic usage.

18 All infants were examined for atopic dermatitis (AD) at their enrolment visit at three months of
19 age and twelve months, using the UK diagnostic criteria-based photographic protocol of the
20 International Study of Asthma and Allergies in Childhood (ISAAC) Phase Two (19). AD severity was
21 determined by the Scoring Atopic Dermatitis (SCORAD) index (20). Skin barrier function was
22 assessed by measuring trans-epidermal water loss (TEWL) using the Biox Aquaflux AF200 (Biox,

London, UK) closed condenser chamber device on unaffected skin of the left volar forearm. Venous blood samples were screened for the 6 commonest *FLG* mutations (TaqMan allelic discrimination assays for mutations R501X, 2282del4, R2447X, S3247X; ABI 7900 HT; Applied Biosystems, Foster City, Calif) or by sizing of fluorescent PCR products on an Applied Biosystems 3130 DNA sequencer (mutations 3673delC and 3702delG). These 6 mutations detect 99% of *FLG* mutation carriers in the UK population.

Ethical approval for the EAT Study was provided by the St Thomas' Hospital Research Ethics Committee (REC Reference 08/H0802/93) and informed consent was obtained from the parents of all children enrolled in the study.

Gut microbiota sample collection, DNA extraction and sequencing

Beginning the 1st of October 2011, consecutively enrolling families (n=359) were asked to provide stool samples (see Fig E1). Stool samples were collected from 288 of enrolled participants with follow up samples requested at six and twelve months of age. Stool samples were retrieved from diapers and transferred to the King's College London Molecular Microbiology Research Laboratory for storage at -80°C until DNA extraction. The cell lysis and DNA extraction protocol utilized both phenol-chloroform liquid-liquid phase separation and bead-beating, followed by -80°C storage (see Supplementary Appendix). To evaluate the assembly of gut microbial communities during infancy, a total of 428 samples were sequenced (of the total 288 subjects providing baseline samples, 218 participants provided baseline samples only, and a subset of 70 individuals provided samples from multiple time points of three, six, and twelve months). The subset of cases and controls (42 in the standard-introduction group and 28 in the early-

introduction group) were selected based on intervention assignment and atopy (either challenge-proven food allergy or AD on examination) status. DNA extracts were amplified using primers targeting the hypervariable V4 (515F-806R) region of the 16S ribosomal RNA gene and performed on the MiSeq instrument (21).

Sequence analysis

The open-source software mothur (22) and DADA2 (23) were used to process the raw sequencing data (see online supplement). Reads were aligned and classified down to the genus level with ribosomal database project naïve Bayesian classifier (24) or species level with Greengenes database v13.8 (25). Operational taxonomic units (OTUs) were defined at 97% similarity using the optclust algorithm (26). We also undertook comparative analyses with 977 samples from the TwinsUK (adults from the same UK region) (27, 28), which were sequenced using the same primers (Accession No. ERP006339 and ERP006342; 9 samples were removed during preprocessing) and TEDDY (large, multi-centre infant cohort) (11) studies. TEDDY dataset samples were selected based on i) exclusive breastfeeding before solid food introduction and ii) longitudinal data (≥ 2 samples before 6 months of age) to permit comparable analyses.

Statistical analysis

The beta-diversity and linear discriminant statistical analyses were performed using R software, with libraries including vegan, ape, and bios2mds. Thetayc distance (theta) metric was used to measure distance between communities (beta diversity) (29), and principal coordinate analysis

1 (PCoA) was performed based on theta distance, unless otherwise indicated. Significance of
2 differences between groups was tested by various methods, including Wilcoxon rank-sum,
3 Kruskal-wallis, and the analysis of molecular variance (AMOVA) (30). Covariates of community
4 variation were evaluated by calculating correlation between indicated PCoA ordination and
5 metadata or genus abundance (envfit function in vegan R package; 10,000 permutations with
6 Bonferroni correction).

Results

Of enrolled families, 80.2% (288 of 359) contributed baseline samples, a subset of which (n=70) also contributed follow-up samples. Within the subset, cases were comprised of twelve participants with allergies to egg (one additionally allergic to peanut), one milk allergy, and one codfish allergy, and thirteen with AD (SCORAD over 15 at either three or twelve months).

Exclusively breastfed infants' gut microbial communities form 3 distinct clusters

To understand the landscape of the gut microbiome during breastfeeding, we analyzed baseline (3-month) samples (n=288, and a mean age of 105 days and interquartile range 100-112 days).

The most abundant genera established in the microbial communities at three months were *Bifidobacterium* (median abundance (MA): 38.75%), *Bacteroides* (MA: 5.43%) and multiple genera belonging to Firmicutes and Proteobacteria (Fig 1, A and B). The gut microbial communities of exclusively breastfed three-month-old infants was diverse, in examining the PCoA with theta distance (29) and Bray-Curtis distance based on genus-level classification (Fig 1, A and see Fig E2, A). k-means clustering (see Fig E2, A and B) revealed that the microbial communities of exclusively breastfed infants clustered into 3 groups, in which cluster with *Bifidobacterium*-rich cluster #1, *Bacteroides*-rich cluster #2 and cluster #3 predominated by *Escherichia/Shigella* and genera belonging to Firmicutes phylum (Fig 1, A and B). Although minor genera, such as *Streptococcus*, *Dorea* and *Rothia*, also contributed to the community variation, the three major genera were sufficient to reproduce clustering, underscoring their importance in community variation (Fig 1, C). Community diversity was lower in cluster #1 than in clusters #2 and #3 (alpha diversity; Shannon index, $p < 0.005$; Wilcoxon-rank sum test) due to the predominance of *Bifidobacterium* (Fig 1, D). Of note, the age of participants did not have a

significant effect on clustering (see Fig E2, C, $p>0.05$). PCoA and k-means clustering based on theta distance calculated from species-level (see Fig E2, D) and *de novo* OTU (97% similarity) abundance (see Fig E2, E) and on Bray-Curtis distance (see Fig E2, A) were highly comparable to the genus-based clustering (Fig 1, B).

Infants' gut microbial communities mature in first year

Data from six- and twelve-month samples ($n=70$ samples per each time point; see methods) were superimposed onto the existing three-month PCoA coordinates (see Fig E3, A). Microbiota from six-month-old infants were similarly dispersed as three-month communities, whereas twelve-month samples were exclusively overlaid on cluster #2, indicating that gut microbial communities mature towards *Bacteroides*-rich communities at 1 year of age, in keeping with the previous literature (7, 10). Interpersonal similarity increased during aging (see Fig E3, B) due to the convergence to *Bacteroides*-rich communities. We then investigated whether the infant gut microbiomes matured towards an adult-like pattern. Since gut microbiomes can vary in adult populations from distinct geographic regions (31), we compared our dataset to a representative dataset of fecal microbiota from adults in a geographic region analogous to our infant cohort, revealing that infants begin to transition towards adult gut microbiomes during six to twelve months of age (see Fig E4). Collectively, these results demonstrated the extensive transition and maturation of gut microbiota during the first year of life.

Early environmental exposures are associated with the infant gut microbiota

Among the environmental exposures that were evaluated, mode of delivery was the most significant factor contributing to microbiome community variation ($p < 0.01$) at 3 months of age (Fig 2, A), with underrepresentation of *Bacteroides* in individuals born by cesarean section (Fig 2, B and C), as shown in previous reports (5, 32, 33). Since the mode of delivery profoundly contributed to microbiome variation, PCoA and clustering were re-analyzed after stratification by delivery methods. Gut microbial communities of vaginally delivered infants also formed 3 clusters, with the same three key genera – *Bifidobacterium*, *Bacteriodes*, and *Escherichia/Shigella* – as the major contributing taxa (see Fig E5, A), indicating contribution of other exposures in defining microbiome communities. Therefore, pair-wise associations between major bacterial genera (mean abundance $> 1\%$) and environmental factors were investigated (see Fig E5, B). Results highlighted additional bacterial genera associated with delivery mode (e.g. *Bifidobacterium*, *Clostridium sensu stricto*, Enterobacteriaceae, and Lachnospiraceae were enriched in infants with cesarean-delivered infants) underscoring the importance of delivery mode in shaping infant gut microbiomes.

Clinical outcomes are linked with specific infant gut microbiota

We next examined potential links between clinical outcomes and the 3-month gut microbiome clusters; however, no statistically significant associations were identified. Therefore, associations between major bacterial genera (mean abundance $> 1\%$) and clinical manifestations were also investigated (Fig, E6, A). Interestingly, SCORAD at 12 months of age was significantly associated with the abundance of *Clostridium sensu stricto* and *Haemophilus* at enrolment. Equally, raised TEWL at 3 months showed a positive association with *Haemophilus*. Raised TEWL at 12 months

was also linked to an increased abundance in *Veillonella* (Fig E6, A). Linear discriminant analysis confirmed higher relative abundances of *Clostridium sensu stricto* at three months of age were observed in infants who had AD at three months and twelve months of age (see Fig E6, B and C). Family history of allergic disease and food allergy/sensitization did not significantly differ according to the microbiome variation or clustering.

The introduction of allergenic foods alters infants' gut microbiota

The EAT Study randomized exclusively breastfed infants to dietary introduction of six allergenic foods ('early introduction', n=28) versus continued exclusive breastfeeding ('standard introduction', n=42) until 6 months of age (17). Therefore, the maturation in the gut microbiota was investigated relative to the timing of introduction of allergenic solids. Participants in the intervention group had a significantly increased Shannon microbial diversity index at six months as compared to their enrolment samples (Fig 3, A, and Fig E7, A). PCoA revealed the gut microbial communities of the early and standard introduction groups were largely indistinguishable (Fig 3, B, P non-significant; AMOVA) at three and twelve months. In contrast, these microbial communities significantly differed at six months of age (Fig 3, B, middle panel. $p < 0.05$; AMOVA), demonstrating the differential maturation trajectories of early versus standard introduction groups at six months of age. Interestingly, early intervention group infants predominantly transitioned along the second principal coordinate (y-axis) from three to six months of age (Fig 3, C). Abundances of *Prevotella* and *Escherichia/Shigella* were positively correlated with the second principal coordinate ($p < 0.05$, and $r^2 = 0.52$ and 0.44 , respectively). Similarly, linear discriminant analysis revealed *Paraprevotella* and various genera belonging to Proteobacteria and Firmicutes

1 phyla were enriched in the early intervention group (see Fig E7, B and D) at six months, but not
2 evident at 12 months of age (see Fig E7, C and E). Infants in both groups moved along with the
3 first principal coordinate (x-axis) and converged towards *Bacteroides*-rich communities during six
4 to twelve months (Fig 3, D). To validate that the introduction of solids altered the gut microbiota
5 in exclusively breastfed infants, we analyzed gut microbiome changes in exclusively breastfed
6 infants after similar dietary changes within the TEDDY cohort (see Methods for detailed criteria)
7 (11) and superimposed their data onto the same PCoA dimensions of (see E8, A). TEDDY study
8 infants with early introduction of solids (n=48, no formula and started solid foods before 150
9 days of life) also transitioned along with the second principal coordinate from 3 months to 6
10 months, whereas TEDDY study infants mimicking the standard introduction group (n=45, no
11 formula and started solid foods after 150 days of life) did not follow the same gut microbiome
12 transition (Fig 3, E).

1 Discussion

2 There remains deep interest in understanding what contribution external factors may contribute
3 to the development of atopic and allergic disorders. This cohort provided a unique opportunity
4 to examine the relationship between the gut microbiome and many factors including dietary
5 intervention, clinical outcomes, and hygiene-related factors. Gut microbiota are widely
6 heterogenous amongst exclusively breastfed three-month-old infants and have lesser abundance
7 of *Bacteroides* species if born by cesarean section. After three months of age, the gut microbiota
8 underwent major transformation in diversity, complexity, and predominance of 'adult-like'
9 bacteria. Babies' gut microbiota diversified when allergenic foods were introduced and matured
10 towards *Bacteroides*-rich communities at twelve months of age, with significant changes
11 observed at a younger age in infants with earlier introduction of allergenic solids beginning at
12 three months of age. While there were significant microbiota-related associations with AD and
13 AD severity at three and 12 months, none of the observed gut microbiota changes were
14 associated with the development of food allergy or sensitization.

15

16 *Study strengths and limitations*

17 Strengths of the EAT study cohort were strictly controlled eligibility of exclusively breastfed
18 infants, population-representative sampling from the general population across England and
19 Wales, and a randomized intervention with allergenic solid introduction between three- and six-
20 months age, rather than an observational study design with no control of food intake (17). The
21 nested observational study within the EAT cohort allowed comparisons of the effects of the

1 dietary intervention on the gut microbiota over the first year of life. We validated these findings
2 in exclusively breastfed infants with comparable introductions of solid foods in the TEDDY cohort,
3 recapitulating these observations in other populations with westernized diets. Detailed
4 phenotyping of this cohort enabled assessment of alterations in infant gut microbiota in
5 association with exposures, such as mode of delivery, antibiotic prescribing, communal child
6 care, sibship size, rural vs urban residence, and pet ownership. Validated scores were used to
7 examine participants for AD, and diagnosis of food allergy was challenge-based, rather than
8 relying on parental report. A limitation of our study is that we were unable to evaluate the gut
9 microbiota before three months of life, which would have allowed an even earlier
10 characterization of microbiota development.

11 *Neonatal environmental exposures may influence gut microbiota at three months of age*

13 Delivery by cesarean had the most profound effects on the shaping of the gut microbiota during
14 early infancy with reduction in *Bacteroides*, similar to other cohorts (5, 11, 32, 33). Colonization
15 with *Bacteroides* in the neonatal period may have considerable health benefits, especially
16 immune maturation (34). *Bacteroides fragilis* has Polysaccharide A molecular moieties which
17 stimulate the development of T-regulatory cells in the lamina propria (35) and samples from
18 human fecal microbiota have found that at least 16 *Bacteroides* taxa contribute to this
19 phenomenon *in vivo* (36). Therefore, the abundance of *Bacteroides* spp. may relate to the nature
20 of early life immunological priming and warrants further investigation. In addition, if these
21 mechanistic studies are fruitful, intervention trials may consider targeting infants born by
22 cesarean to try and alter *Bacteroides* relative abundances and/or *Bacteroides* diversity (7).

Although we did not find other exposures contributing to the heterogeneity of microbiome at three months of infancy, there may be mixed effects of multiple factors and/or hidden factors which were not included in our metadata; therefore, an optimized model for discovering associated factors might be necessary. With the current dataset, we trained a random forest classifier with metadata to predict 3 clusters at three months. While the classifier did not perform well (out-of-box error rate > 40%, data not shown), it calculated maternal age at birth as a second important factor for classifying clusters, suggesting the importance of maternal/household factors in microbiome development at infancy. Similarly, studies have reported the long-term colonization of vertically transmitted microbes (12).

Diet contributes to evolution of the infant gut microbiota

Infants' diet also had significant effects on the shaping of the gut microbiota during early infancy. Several reports have demonstrated differential microbiome colonization in breastfed versus formula-fed infants, with a higher prevalence of *Bifidobacterium* spp. (*B. longum* and *B. breve*) and *Lactobacillus* in breastfed infants (8, 9, 11, 14). In addition, the introduction of solid foods has been proposed as a key contributor to the maturation of the gut microbiota. For example, a Scandinavian observational study concluded that the gut microbiota matured after the cessation of breastfeeding (8).

Since studies randomizing dietary interventions in infants are very unusual, this study offered a unique opportunity to study the effects of introducing allergenic solids on the gut microbiome. In the EAT study, early peanut and egg introduction, if consumed in sufficient quantity, was

1 shown to protect against the development of peanut and egg allergies between one and three
2 years of age (17, 18). We have demonstrated that the early introduction of allergenic foods
3 alongside ongoing breastfeeding between three and six months of age led to an increase in
4 overall gut microbiota Shannon diversity, in particular promoting an influx of various microbes
5 including *Prevotellaceae* and *Escherichia/Shigella*. Interestingly, presence of *Prevotella* has been
6 shown to be associated with high-fiber diet including remote villages with less frequent chronic
7 inflammatory disorders (37-40). Additionally, preceding studies have noted that the cessation of
8 breastfeeding may promote further diversification of infants' gut microbiota(8), whereas our
9 results showed that the introduction of a variety of allergenic foods in the setting of exclusive
10 breastfeeding increased diversity of the gut microbiota. Of note, six-month gut microbiota
11 Shannon diversities were not statistically significantly different between the early and standard
12 introduction groups, indicating that the increased microbial community diversity after early
13 introduction of allergenic foods was comparable to exclusively breast-fed infants (Fig E7, A).
14 However, the combined findings from this study demonstrate that early introduction of solid
15 foods elicits considerable and consistent changes in gut microbial composition and diversity from
16 three to six months of infancy. Therefore, controlled dietary interventions in infancy may
17 promote selective colonization of desired microbes and potentially exerts health and
18 developmental benefit to the host. However, we found no evidence that the observed *changes*
19 in the gut microbiota driven by the early introduction of allergenic foods impact on the
20 development of AD.

21 In conclusion, we confirmed that cesarean delivery is associated with reduced gut microbiota
22 diversity at three months. The early introduction of solids into infants' diets accelerates

- 1 maturation of microbiota diversity as well as increases the relative abundances of *Prevotella* and
- 2 *Escherichia/Shigella*. This manipulation of the gut microbiota in early life towards diversity and
- 3 maturity, in particular in those delivered by cesarean section or at risk of allergy and atopic
- 4 diseases, should be examined in future studies to examine potential wider health benefits.

1 **Acknowledgements**

2 We thank the parents and children for taking part in this trial; the members of the trial steering
3 committee (Graham Roberts [chair], David Strachan [vice chair], Mary Fewtrell, Christine
4 Edwards, David Reading, Ian Kimber, Anne Greenough, Andy Grieve, Mary Feeney, Kate
5 Grimshaw, Judy More, Debbie Palmer, Carina Venter, and Rebecca Knibb) for contributions to
6 the study design; Monica Basting and Gemma Deutsch for project management; Helen Fisher,
7 Una O'Dwyer-Leeson, Amy Nixon, Louise Coverdale, and Muhsinah Adam for nursing support;
8 Alicia Parr for dietetic support; George Du Toit and Susan Chan for assistance with medical
9 supervision; Jenna Heath and Kathryn Hersee for play-specialist support; and Joelle Buck, Sarah
10 Hardy, Elizabeth Kendall, and Shuhana Begum, of the Food Standards Agency, for their
11 commitment to the trial.

Data Access

All sequencing data were deposited and are available at the NCBI Sequence Read Archive (SRA) under BioProject number PRJNA597342.

References

1. Zeissig S, Blumberg RS. Life at the beginning: perturbation of the microbiota by antibiotics in early life and its role in health and disease. *Nat Immunol*. 2014;15(4):307-10.
2. Arrieta MC, Stiemsma LT, Dimitriu PA, Thorson L, Russell S, Yurist-Doutsch S, et al. Early infancy microbial and metabolic alterations affect risk of childhood asthma. *Sci Transl Med*. 2015;7(307):307ra152.
3. Fujimura KE, Sitarik AR, Havstad S, Lin DL, Levan S, Fadrosch D, et al. Neonatal gut microbiota associates with childhood multisensitized atopy and T cell differentiation. *Nat Med*. 2016;22(10):1187-91.
4. Stokholm J, Blaser MJ, Thorsen J, Rasmussen MA, Waage J, Vinding RK, et al. Maturation of the gut microbiome and risk of asthma in childhood. *Nat Commun*. 2018;9(1):141.
5. Dominguez-Bello MG, Costello EK, Contreras M, Magris M, Hidalgo G, Fierer N, et al. Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proc Natl Acad Sci U S A*. 2010;107(26):11971-5.
6. Stokholm J, Thorsen J, Chawes BL, Schjorring S, Krogfelt KA, Bonnelykke K, et al. Cesarean section changes neonatal gut colonization. *J Allergy Clin Immunol*. 2016;138(3):881-9 e2.
7. Bokulich NA, Chung J, Battaglia T, Henderson N, Jay M, Li H, et al. Antibiotics, birth mode, and diet shape microbiome maturation during early life. *Sci Transl Med*. 2016;8(343):343ra82.
8. Backhed F, Roswall J, Peng Y, Feng Q, Jia H, Kovatcheva-Datchary P, et al. Dynamics and Stabilization of the Human Gut Microbiome during the First Year of Life. *Cell Host Microbe*. 2015;17(6):852.
9. Baumann-Dudenhoeffer AM, D'Souza AW, Tarr PI, Warner BB, Dantas G. Infant diet and maternal gestational weight gain predict early metabolic maturation of gut microbiomes. *Nat Med*. 2018.
10. Yassour M, Vatanen T, Siljander H, Hamalainen AM, Harkonen T, Ryhanen SJ, et al. Natural history of the infant gut microbiome and impact of antibiotic treatment on bacterial strain diversity and stability. *Sci Transl Med*. 2016;8(343):343ra81.
11. Stewart CJ, Ajami NJ, O'Brien JL, Hutchinson DS, Smith DP, Wong MC, et al. Temporal development of the gut microbiome in early childhood from the TEDDY study. *Nature*. 2018;562(7728):583-8.
12. Shao Y, Forster SC, Tsaliki E, Vervier K, Strang A, Simpson N, et al. Stunted microbiota and opportunistic pathogen colonization in caesarean-section birth. *Nature*. 2019;574(7776):117-21.
13. Penders J, Thijs C, Vink C, Stelma FF, Snijders B, Kummeling I, et al. Factors influencing the composition of the intestinal microbiota in early infancy. *Pediatrics*. 2006;118(2):511-21.
14. Bezirtzoglou E, Tsiotsias A, Welling GW. Microbiota profile in feces of breast- and formula-fed newborns by using fluorescence in situ hybridization (FISH). *Anaerobe*. 2011;17(6):478-82.
15. Palmer C, Bik EM, DiGiulio DB, Relman DA, Brown PO. Development of the human infant intestinal microbiota. *PLoS Biol*. 2007;5(7):e177.

16. Koenig JE, Spor A, Scalfone N, Fricker AD, Stombaugh J, Knight R, et al. Succession of microbial consortia in the developing infant gut microbiome. *Proc Natl Acad Sci U S A*. 2011;108 Suppl 1:4578-85.
17. Perkin MR, Logan K, Marrs T, Radulovic S, Craven J, Flohr C, et al. Enquiring About Tolerance (EAT) study: Feasibility of an early allergenic food introduction regimen. *J Allergy Clin Immunol*. 2016;137(5):1477-86 e8.
18. Perkin MR, Logan K, Tseng A, Raji B, Ayis S, Peacock J, et al. Randomized Trial of Introduction of Allergenic Foods in Breast-Fed Infants. *N Engl J Med*. 2016;374(18):1733-43.
19. Weiland SK, Bjorksten B, Brunekreef B, Cookson WO, von Mutius E, Strachan DP, et al. Phase II of the International Study of Asthma and Allergies in Childhood (ISAAC II): rationale and methods. *Eur Respir J*. 2004;24(3):406-12.
20. Kunz B, Oranje AP, Labreze L, Stalder JF, Ring J, Taieb A. Clinical validation and guidelines for the SCORAD index: consensus report of the European Task Force on Atopic Dermatitis. *Dermatology*. 1997;195(1):10-9.
21. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, et al. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc Natl Acad Sci U S A*. 2011;108 Suppl 1:4516-22.
22. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol*. 2009;75(23):7537-41.
23. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods*. 2016;13(7):581-3.
24. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol*. 2007;73(16):5261-7.
25. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol*. 2006;72(7):5069-72.
26. Westcott SL, Schloss PD. OptiClust, an Improved Method for Assigning Amplicon-Based Sequence Data to Operational Taxonomic Units. *mSphere*. 2017;2(2).
27. Goodrich JK, Waters JL, Poole AC, Sutter JL, Koren O, Blekhman R, et al. Human genetics shape the gut microbiome. *Cell*. 2014;159(4):789-99.
28. Jackson MA, Verdi S, Maxan ME, Shin CM, Zierer J, Bowyer RCE, et al. Gut microbiota associations with common diseases and prescription medications in a population-based cohort. *Nat Commun*. 2018;9(1):2655.
29. Yue JC, Clayton MK. A similarity measure based on species proportions. *Communications in Statistics-Theory and Methods*. 2005;34(11):2123-31.
30. Excoffier L, Smouse PE, Quattro JM. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*. 1992;131(2):479-91.
31. Clemente JC, Pehrsson EC, Blaser MJ, Sandhu K, Gao Z, Wang B, et al. The microbiome of uncontacted Amerindians. *Sci Adv*. 2015;1(3).

32. Rutayisire E, Huang K, Liu Y, Tao F. The mode of delivery affects the diversity and colonization pattern of the gut microbiota during the first year of infants' life: a systematic review. *BMC Gastroenterol.* 2016;16(1):86.
33. Jakobsson HE, Abrahamsson TR, Jenmalm MC, Harris K, Quince C, Jernberg C, et al. Decreased gut microbiota diversity, delayed Bacteroidetes colonisation and reduced Th1 responses in infants delivered by caesarean section. *Gut.* 2014;63(4):559-66.
34. Kelly D, Campbell JI, King TP, Grant G, Jansson EA, Coutts AG, et al. Commensal anaerobic gut bacteria attenuate inflammation by regulating nuclear-cytoplasmic shuttling of PPAR-gamma and RelA. *Nat Immunol.* 2004;5(1):104-12.
35. Round JL, Mazmanian SK. Inducible Foxp3⁺ regulatory T-cell development by a commensal bacterium of the intestinal microbiota. *Proc Natl Acad Sci U S A.* 2010;107(27):12204-9.
36. Atarashi K, Tanoue T, Shima T, Imaoka A, Kuwahara T, Momose Y, et al. Induction of colonic regulatory T cells by indigenous Clostridium species. *Science.* 2011;331(6015):337-41.
37. Yatsunenkov T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, et al. Human gut microbiome viewed across age and geography. *Nature.* 2012;486(7402):222-7.
38. De Filippo C, Cavalieri D, Di Paola M, Ramazzotti M, Poullet JB, Massart S, et al. Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. *Proc Natl Acad Sci U S A.* 2010;107(33):14691-6.
39. Schnorr SL, Candela M, Rampelli S, Centanni M, Consolandi C, Basaglia G, et al. Gut microbiome of the Hadza hunter-gatherers. *Nat Commun.* 2014;5:3654.
40. Marungruang N, Tovar J, Bjorck I, Hallenius FF. Improvement in cardiometabolic risk markers following a multifunctional diet is associated with gut microbial taxa in healthy overweight and obese subjects. *Eur J Nutr.* 2018;57(8):2927-36.

List of figure captions**FIG 1. Gut microbiota characteristics at three months of age**

A, Principal coordinate analysis (PCoA) of the gut microbiome at 3 months. Pair-wise distances (theta distance) among all samples were calculated and two major axes (PC1 and PC2) from the multi-dimensional distance space were calculated and depicted on a scatter plot. Colors indicate different clusters, according to k-means clustering. Arrows and letters indicate specific genera significantly correlated with PCoA ordination ($p < 0.05$, lengths of arrows are proportion to R^2 , calculated from EnvFit function from R). Corresponding genera for letters are showing on the right panel of B.

B, Averaged stacked bar chart of infants' microbiome within each cluster. Genera that are significantly correlated with PCoA ordination are depicted as colored bar (legend on right) and others are merged as 'Other bacteria'.

C, Triangle plot showing relative abundances of the three key genera (*Bifidobacterium*, *Bacteroides*, and *Escherichia/Shigella*) in the gut microbiota of baseline samples. The colors indicate clusters (same as Fig 1A).

D, Boxplot of bacterial community diversity (Shannon index) differences according to microbiome community clusters. Shannon diversity of Cluster #1 community is significantly lower than Cluster 2 and 3. (** $p < 0.01$, *** $p < 0.001$; Wilcoxon rank-sum test, after Kruskal-wallis test)

FIG 2. Evaluation of associations between gut microbiota characteristics and cesarean delivery

A, Association analysis of clinical, hygiene and life-style factors ('covariates') with microbiome variations. Among total of 48 covariates tested, 11 covariates with $r^2 > 0.01$ are plotted here.

Each bar indicates the amount of variance explained by each covariate, calculated by EnvFit function from R. (** $p < 0.01$; Bonferroni corrected) After correction for multiple comparison, mode of delivery remained as a significant factor that associated to microbiome variation.

B, PCoA of microbiomes at 3 months with delivery mode delineated. Color indicates cluster (same as Fig 1, A) and the shape indicates delivery mode. Arrows demonstrate direction of each covariate in the ordination space. Infants born by cesarean section tend to cluster on the left portion of scatter plot.

C, Boxplot showing the difference of relative abundances of *Bacteroides* by mode of delivery. *Bacteroides* are largely missing in infants born by cesarean section. (** $p < 0.001$; Wilcoxon rank-sum test)

FIG 3. Impact of allergenic solid introduction on infants' gut microbiota

A, Boxplot comparing Shannon diversity changes amongst participants' longitudinal samples according to randomized allocation to continued exclusive breastfeeding (standard introduction group) or the introduction of allergenic solids (early introduction group) (* $p < 0.05$; Wilcoxon rank-sum test).

B, PCoA scatter plot demonstrating longitudinal transition from 3 to 6 month of age. Grey arrows connect samples from the same individuals. Yellow and purple arrows on the sides indicate the average shift of the microbiota in each PCoA axis.

C, PCoA scatter plot demonstrating longitudinal transition from 6 to 12 month of age. Grey arrows connect samples from the same individual. Black arrows on the sides indicate average shift of the microbiota in each PCoA axis.

D, PCoA plot showing longitudinal transition of the gut microbiome from 3 to 12 months of age in the different dietary intervention groups. At 6 months, the microbiota of early introduction is significantly different from standard introduction ($p < 0.05$ by AMOVA).

E, Boxplot showing changes of microbiome from 3 to 6 months, in different dietary intervention groups and cohorts (EAT and TEDDY). Early introduction in both studies led to an increase along the second principal coordinate axis (* $p < 0.05$ and ** $p < 0.01$; paired Wilcoxon rank-sum test).

FIG 1

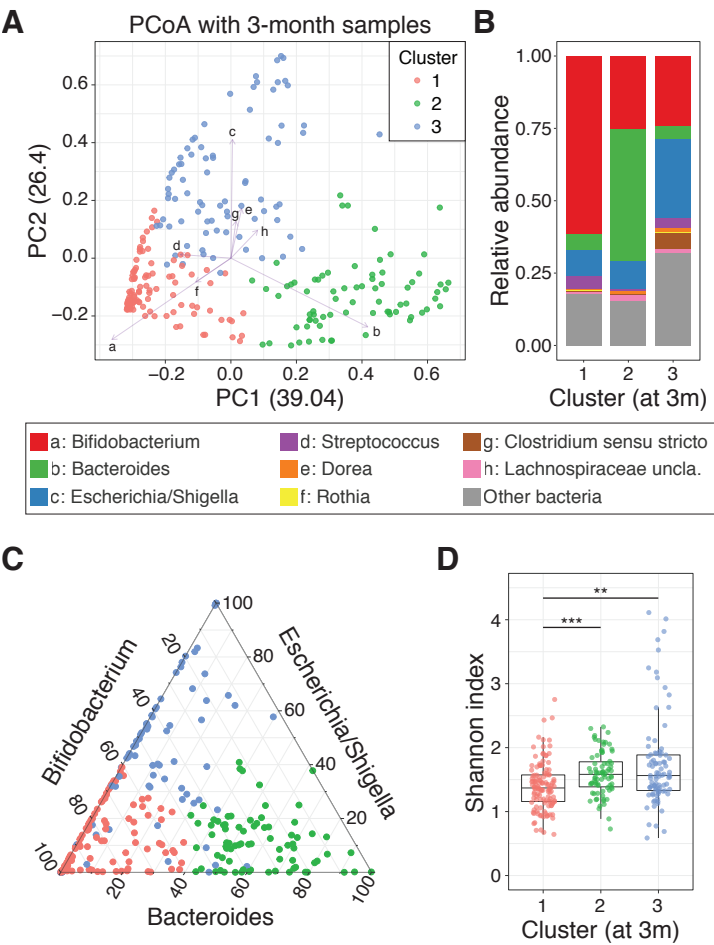


FIG 2

A Association of clinical, hygiene and life-style factors with microbiome at 3m

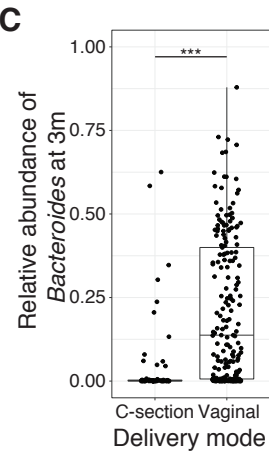
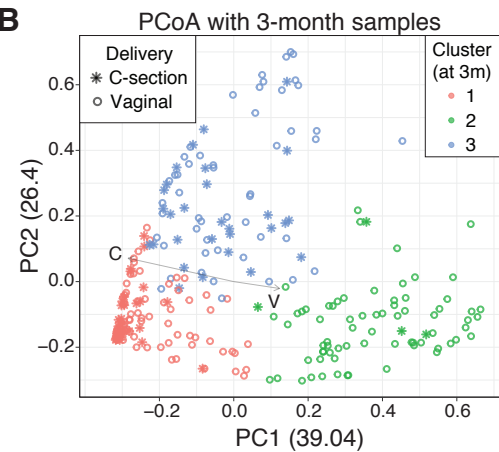
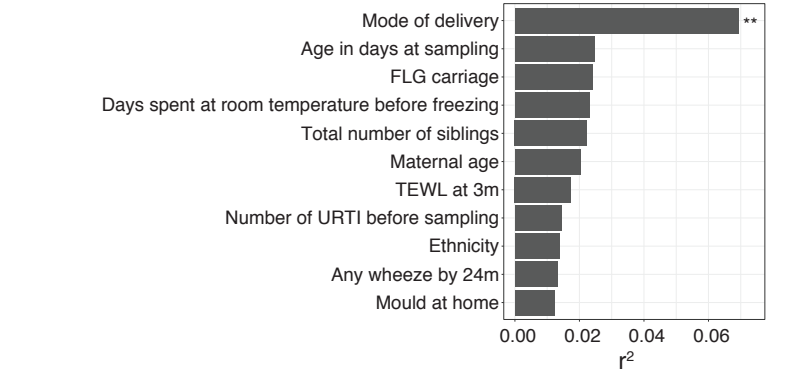
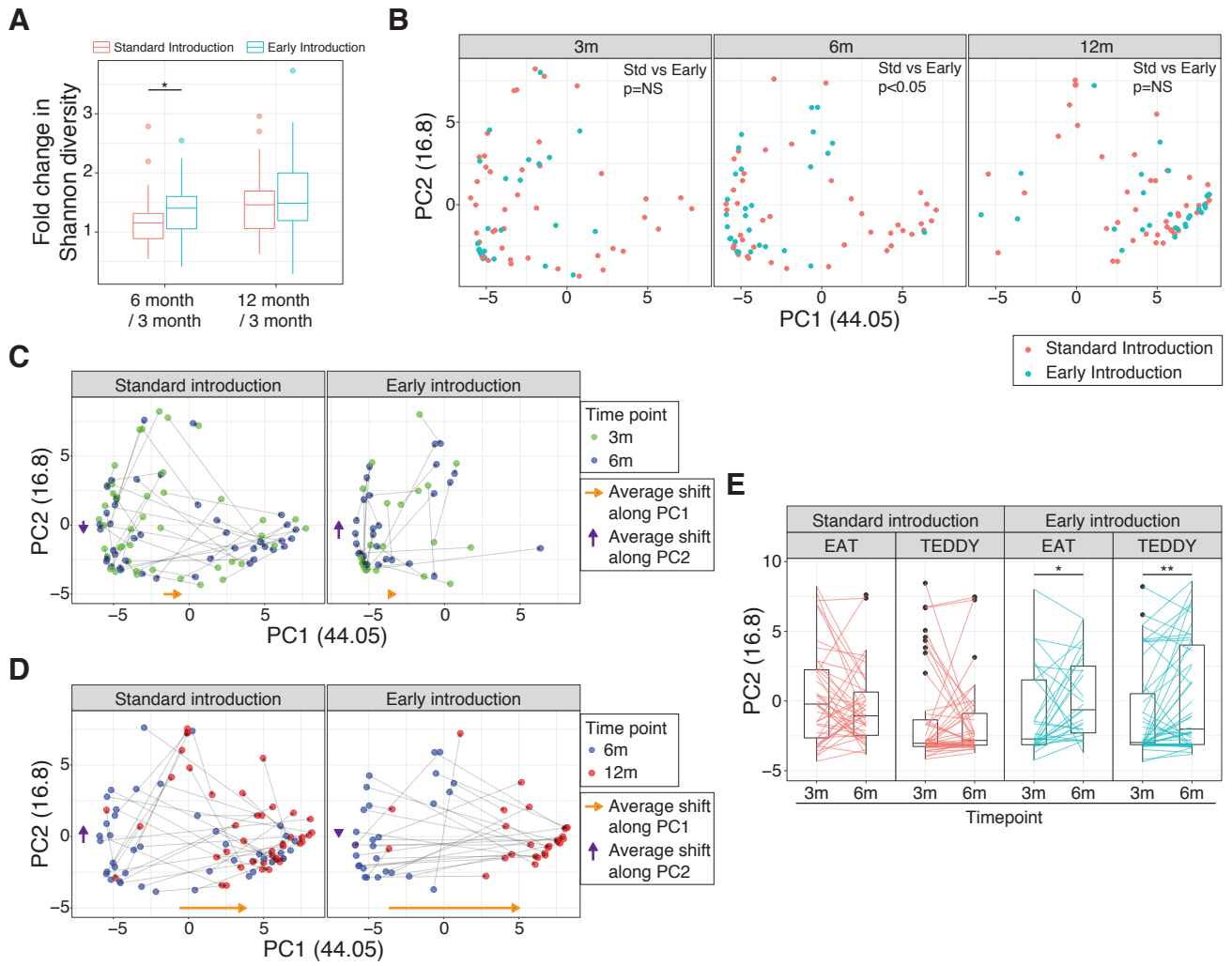


FIG 3



Gut microbiota development during infancy: impact of introducing allergenic foods

Online Materials

FIG E1. Participant flow diagram.

Of the potential participants (n=359) in the nested study, 288 participants provided a baseline (3 months) fecal sample. For the subset of case-control longitudinal sampling (n=70), participants provided additional fecal samples at 6 and 12 months.

FIG E2. Patterns of gut microbiota of baseline fecal samples

A, PCoA of the gut microbiome at 3 months. Pair-wise distances (Bray-Curtis distance) among all samples were calculated and two major axes (PC1 and PC2) from the multi-dimensional distance space were calculated and depicted on a scatter plot. Colors indicate different clusters, according to k-means clustering.

B, Density distribution of three-month-old infants' microbiome based on PCoA (same as Fig 1A). There are three high-density peaks pronounced, suggesting 3 clusters of microbial community populations.

C, Evaluation of optimal clustering using 'within sum of squares' by each k (number of clusters). Based on the graphical visualization of within sum of squares or the 'elbow method', k=3 (3 clusters) is the optimal number of clusters. Increasing the number of clusters beyond 3 (k>3) results in overfitting, whereas k<3 is not as effective in minimizing intra-cluster variation.

D, Distribution of age (in days) of the infants when their 3 months samples were acquired, grouped by each cluster (p=NS; Kruskal-Wallis test).

E, PCoA of microbiome communities at 3 months with species-level classification. The colors indicate clusters.

F, PCoA of microbiome communities at 3 months with 97% *de novo* OTU level classification. The colors indicate clusters.

FIG E3. Longitudinal changes of EAT gut microbiome

A, Gut microbiota for 6 and 12 months samples were overlaid on the PCoA ordination of the microbiota of three-months samples (same as Fig 1A). The colors indicate clustering and age groups. At twelve months of age the microbiota largely converge to cluster #2 of the three-months gut microbiota.

B, Boxplot showing changes of interpersonal dissimilarity at different time points. All-to-all theta distances within each age group were calculated and plotted. Lower value (close to zero) indicates communities are more similar.

FIG E4. Microbiota differences between infants and adults

A, Stacked bar charts for relative abundances of major genera in all samples, from both EAT (3, 6 and 12 months) and TwinsUK (adults) cohorts. Each bar indicates an individual.

B, Schematic diagram describing differences of core/non-core genera between three-month-old infants and adults. There are 5 core genera (genera exist more than 95% of individual,

Bifidobacterium, *Bacteroides*, *Streptococcus*, *Escherichia/Shigella* and *Veillonella*) from three-month-old infants. Among them, 4 genera (except *Veillonella*) are also core in adults.

C, Boxplot showing relative changes of shared and adult-specific core microbes (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; Wilcoxon rank-sum test after Kruskal-wallis test). At 12 months, relative abundances of adult-specific core genera are increased and shared cores are decreased, compared to 3 months.

D, PCoA plot showing relative changes of gut microbiome during maturation. 12 months samples clustered more closely with adults.

A, PCoA of baseline gut microbiota of vaginally born infants. Different colors indicate clusters and arrows indicate specific genera significantly correlated with PCoA ordination ($p < 0.05$, lengths of arrows are proportional to R^2 (calculated by EnvFit in R)).

B, Heatmap showing associations between environmental exposures and microbial composition. Darker colors indicate lower P values and red marks indicate associations with FDR-corrected P values lower than 0.05.

C. Mean relative abundance differences of selected taxa from B (vaginal vs CS).

D. Mean relative abundance differences of selected taxa from B (Antibiotics before sampling).

FIG E6. Association of clinical measurements with microbial compositions.

A, Heatmap showing associations between various clinical measurements and microbial composition. Darker colors indicate lower P values and red marks indicate associations with FDR-corrected P values lower than 0.05.

B, Discriminatory taxa from three-month gut microbiota in infants with or without eczema at three months of age (LDA score: Linear discriminant analysis score). LDA score is calculated by LEfSe(1).

C, Discriminatory taxa from three-month gut microbiota in infants with or without eczema at twelve months of age. LDA score is calculated by LEfSe.

D. Mean relative abundance differences of selected taxa from B (with or without eczema at three months).

E. Mean relative abundance differences of selected taxa from C (with or without eczema at twelve months).

FIG E7. Differential microbiota dynamics by early food introduction in EAT cohort

A, Boxplot comparing Shannon diversity changes amongst participants' longitudinal samples according to randomized allocation to continued exclusive breastfeeding (standard introduction group) or the introduction of allergenic solids (early introduction group). (* $p < 0.05$; paired Wilcoxon rank-sum test)

B, Discriminatory taxa from six-month gut microbiota of infants in standard versus early introduction (LDA score: Linear discriminant analysis score). LDA score is calculated by LEfSe.

C, Discriminatory taxa from twelve-month gut microbiota of infants in standard versus early introduction (LDA score: Linear discriminant analysis score). LDA score is calculated by LEfSe.

D, Mean relative abundance differences of selected taxa from B (standard versus early introduction).

E, Mean relative abundance differences of selected taxa from C (standard versus early introduction).

FIG E8. Differential microbiota dynamics by early food introduction in TEDDY cohort

A, PCoA scatter plot demonstrating longitudinal transition from 3 to 6 months of age of defined samples from the TEDDY cohort. From the TEDDY dataset samples from infants who were i) exclusively breastfed before solid food introduction and ii) longitudinal data (>2 sampling before 6 months of age) were selected. Grey arrows connect samples from the same individual. Yellow and purple arrows on the sides indicate the average shift of the microbiota in the PCoA axes.

Supplementary appendix

Supplementary methods

Chemical lysis was undertaken by adding 200µls of Guanidinium thiocyanate–EDTA–sarkosyl and 900µls of phosphate-buffered saline to the samples before they underwent bead-beating. Cell disruption was undertaken using Qiagen stainless steel beads and tungsten carbide beads on a Fastprep-24 Instrument (MP Biomedicals Europe, Illkirch, France) running at 6.5meters/second for 45 seconds. Two cycles of thermolysis followed by alternating incubation at 90°C and -20°C for 10 minutes each, before cell debris was pelleted by centrifugation at 13,000 g for 10 minutes. Supernatant was transferred to a fresh microfuge tube, where it was inverted with 140µl of 5 molar sodium chloride and 374µl 40% polyethylene glycol and precipitated for one hour at 4°C. DNA was pelleted by centrifugation at 13,000 g for 10 min and resuspended in 500µl of sterile distilled water. 300µl Phenol/chloroform (1:1) was added and samples were vortexed before centrifugation at 13,000 g for 5 minutes. The upper phase was then transferred to a fresh microfuge tube. Total DNA was then precipitated by the addition of an equal volume of isopropanol and 0.1 volume of 10 molar ammonium acetate and stored at -20°C for one hour. DNA was pelleted by centrifugation at 13,000 g for 10 minutes. Pelleted DNA was washed in 70% ethanol, dried and resuspended in 30µl of sterile distilled water (2, 3).

DNA extracts were amplified using universal bacterial primers targeting the 16S ribosomal RNA gene at hypervariable region V4 (515F-806R), which are tailed with sequences to incorporate Illumina (San Diego, CA) adapters and indexing barcodes (4). Sequencing was performed on the MiSeq instrument using version 2 chemistry and 250 cycles, stratifying the amplicon samples according to three, six and twelve month time points between each plate.

References

1. Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, et al. Metagenomic biomarker discovery and explanation. *Genome Biol.* 2011;12(6):R60.
2. Green GL, Brostoff J, Hudspith B, Michael M, Mylonaki M, Rayment N, et al. Molecular characterization of the bacteria adherent to human colorectal mucosa. *J Appl Microbiol.* 2006;100(3):460-9.
3. Rogers GB, Cuthbertson L, Hoffman LR, Wing PA, Pope C, Hooftman DA, et al. Reducing bias in bacterial community analysis of lower respiratory infections. *ISME J.* 2013;7(4):697-706.
4. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, et al. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc Natl Acad Sci U S A.* 2011;108 Suppl 1:4516-22.

Gut microbiota evolution during infancy: impact of introducing allergenic foods

Online Materials

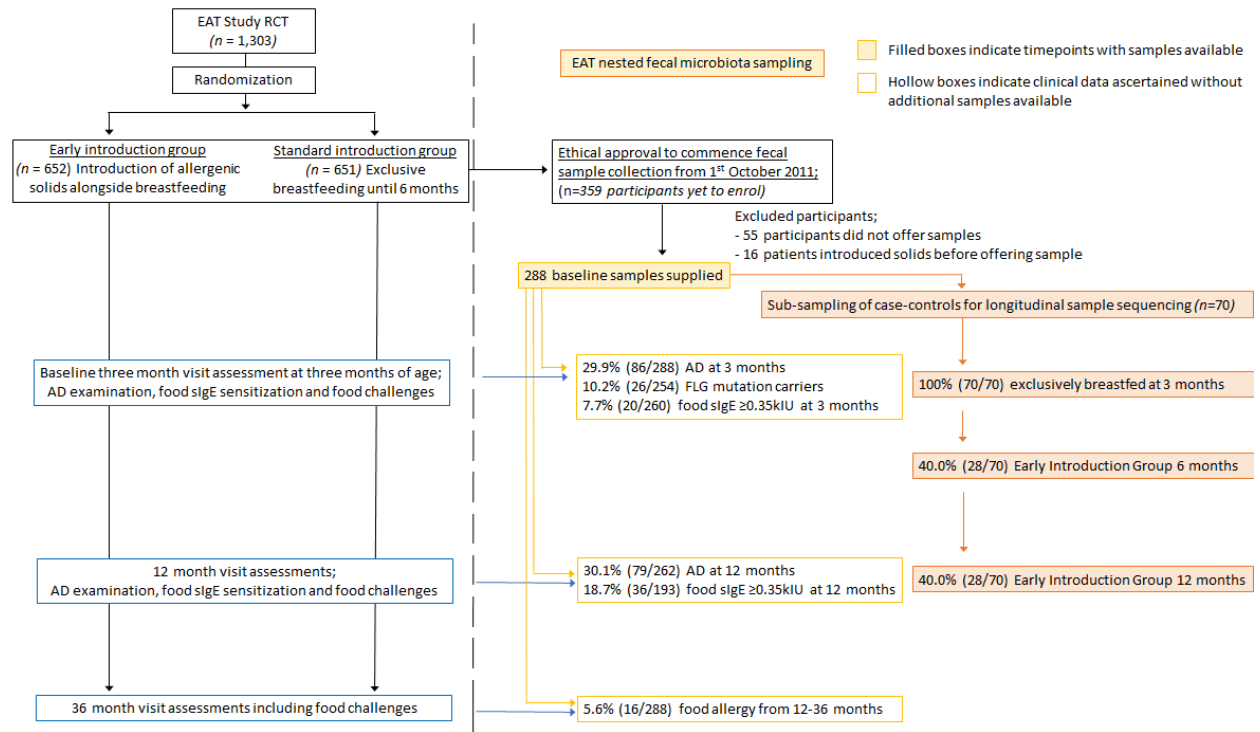


FIG E1. Participant flow diagram.

Of the potential participants (n=359) in the nested study, 288 participants provided a baseline (3 months) fecal sample. For the subset of case-control longitudinal sampling (n=70), participants provided additional fecal samples at 6 and 12 months.

FIG E2

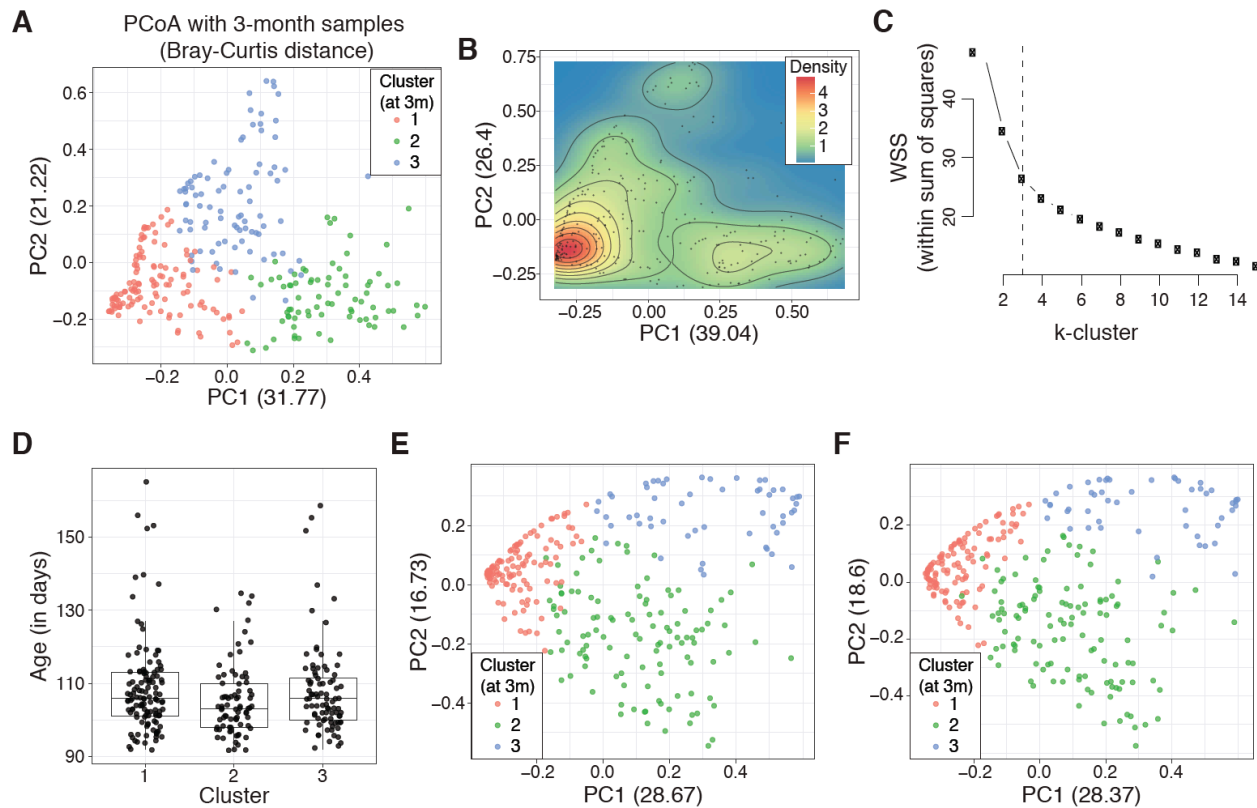


FIG E2. Patterns of gut microbiota of baseline fecal samples

A, PCoA of the gut microbiome at 3 months. Pair-wise distances (Bray-Curtis distance) among all samples were calculated and two major axes (PC1 and PC2) from the multi-dimensional distance space were calculated and depicted on a scatter plot. Colors indicate different clusters, according to k-means clustering.

B, Density distribution of three-month-old infants' microbiome based on PCoA (same as Fig 1A). There are three high-density peaks pronounced, suggesting 3 clusters of microbial community populations.

C, Evaluation of optimal clustering using 'within sum of squares' by each k (number of clusters). Based on the graphical visualization of within sum of squares or the 'elbow method', k=3 (3 clusters) is the optimal number of clusters. Increasing the number of clusters beyond 3 (k>3) results in overfitting, whereas k<3 is not as effective in minimizing intra-cluster variation.

D, Distribution of age (in days) of the infants when their 3 months samples were acquired, grouped by each cluster (p=NS; Kruskal-Wallis test).

E, PCoA of microbiome communities at 3 months with species-level classification. The colors indicate clusters.

F, PCoA of microbiome communities at 3 months with 97% *de novo* OTU level classification. The colors indicate clusters.

FIG E3

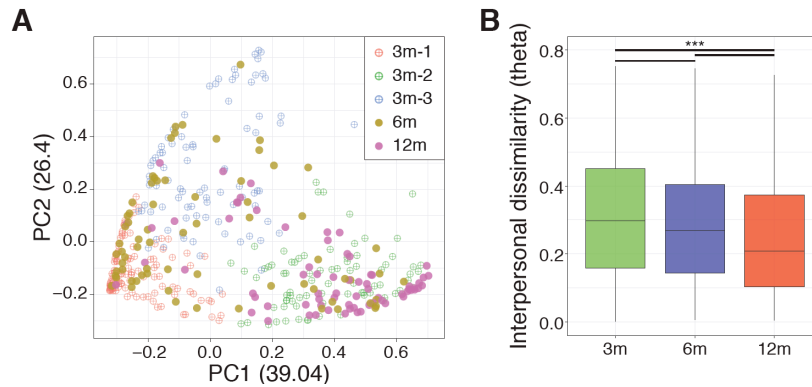


FIG E3. Longitudinal changes of EAT gut microbiome

A, Gut microbiota for 6 and 12 months samples were overlaid on the PCoA ordination of the microbiota of three-months samples (same as Fig 1A). The colors indicate clustering and age groups. At twelve months of age the microbiota largely converge to cluster #2 of the three-months gut microbiota.

B, Boxplot showing changes of interpersonal dissimilarity at different time points. All-to-all theta distances within each age group were calculated and plotted. Lower value (close to zero) indicates communities are more similar.

FIG E4

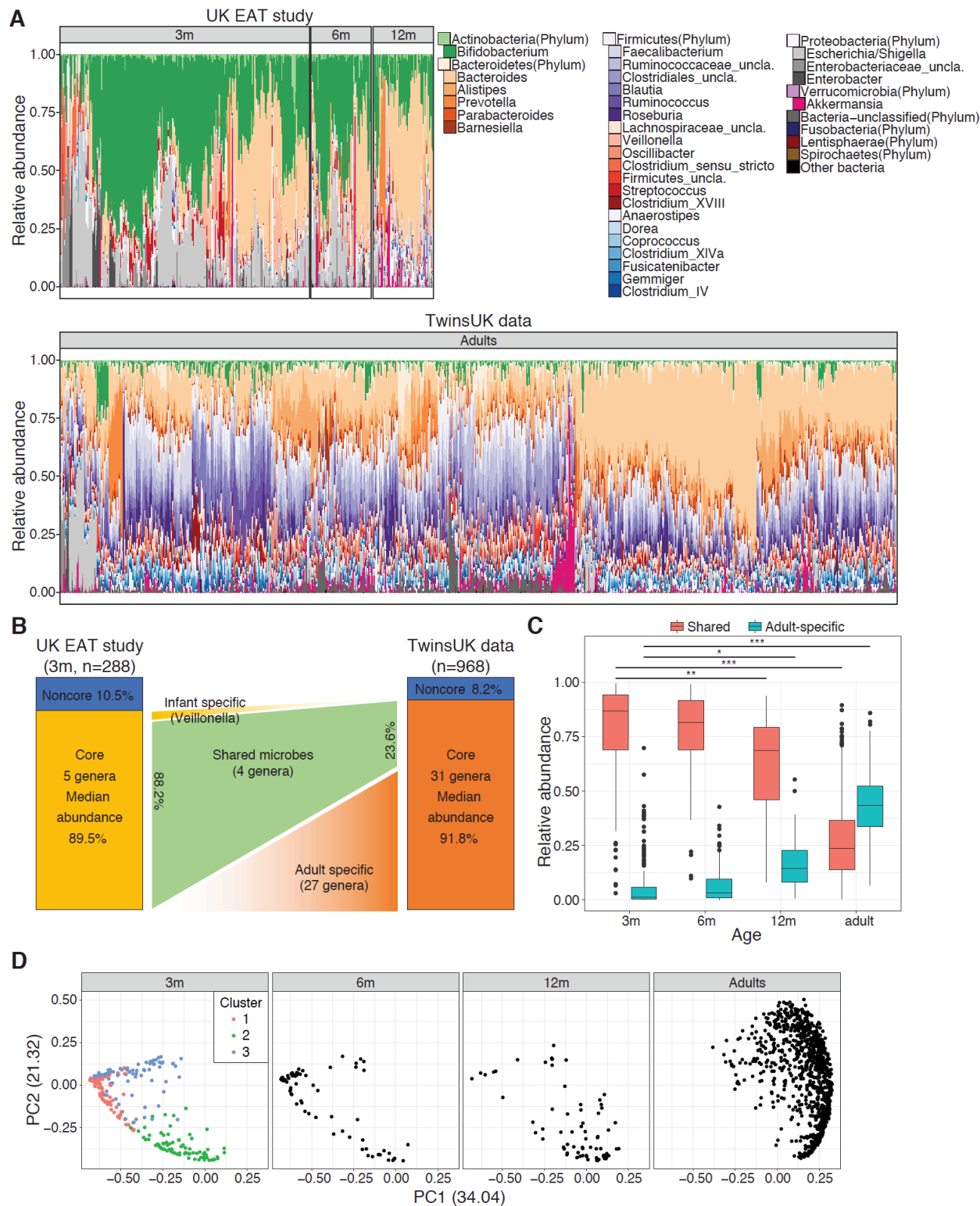


FIG E4. Microbiota differences between infants and adults

A, Stacked bar charts for relative abundances of major genera in all samples, from both EAT (3, 6 and 12 months) and TwinsUK (adults) cohorts. Each bar indicates an individual.

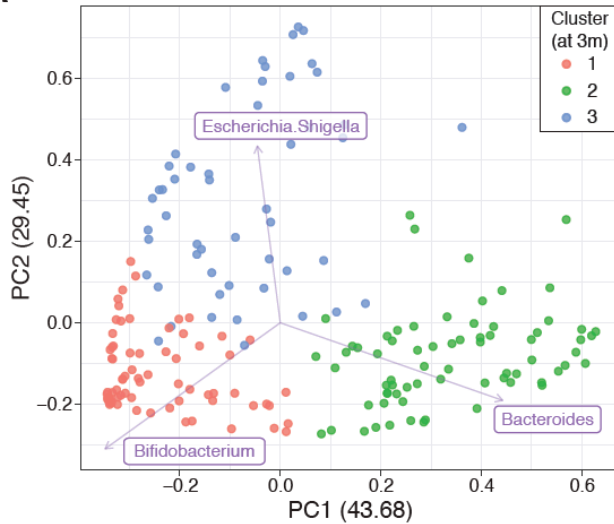
B, Schematic diagram describing differences of core/non-core genera between three-month-old infants and adults. There are 5 core genera (genera exist more than 95% of individual, *Bifidobacterium*, *Bacteroides*, *Streptococcus*, *Escherichia/Shigella* and *Veillonella*) from three-month-old infants. Among them, 4 genera (except *Veillonella*) are also core in adults.

C, Boxplot showing relative changes of shared and adult-specific core microbes (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; Wilcoxon rank-sum test after Kruskal-wallis test). At 12 months, relative abundances of adult-specific core genera are increased and shared cores are decreased, compared to 3 months.

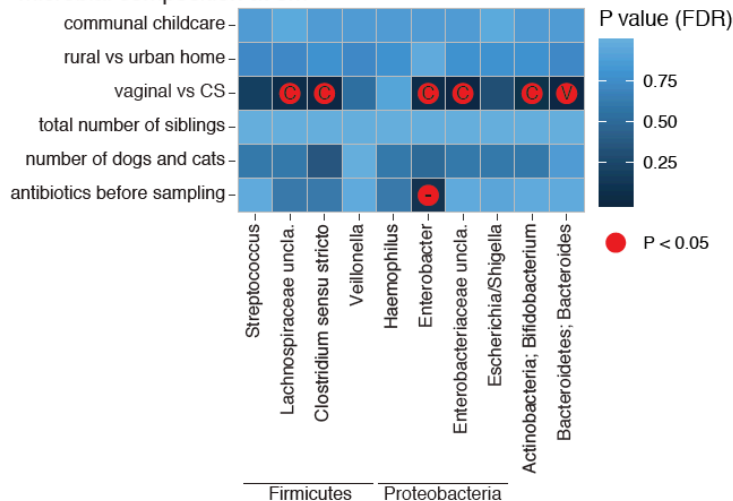
D, PCoA plot showing relative changes of gut microbiome during maturation. 12 months samples clustered more closely with adults.

FIG E5

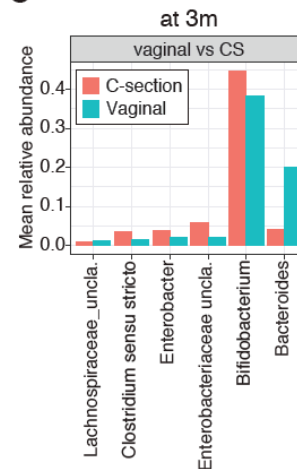
A



B Associations between environmental exposures and microbial composition at 3m



C



D

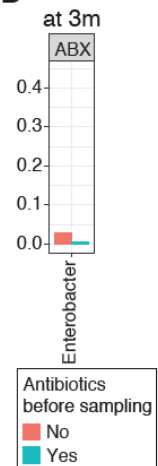


FIG E5. Role environmental factors on microbial compositions.

A, PCoA of baseline gut microbiota of vaginally born infants. Different colors indicate clusters and arrows indicate specific genera significantly correlated with PCoA ordination ($p < 0.05$, lengths of arrows are proportional to R^2 (calculated by EnvFit in R)).

B, Heatmap showing associations between environmental exposures and microbial composition. Darker colors indicate lower P values and red marks indicate associations with FDR-corrected P values lower than 0.05.

C. Mean relative abundance differences of selected taxa from B (vaginal vs CS).

D. Mean relative abundance differences of selected taxa from B (Antibiotics before sampling).

FIG E6

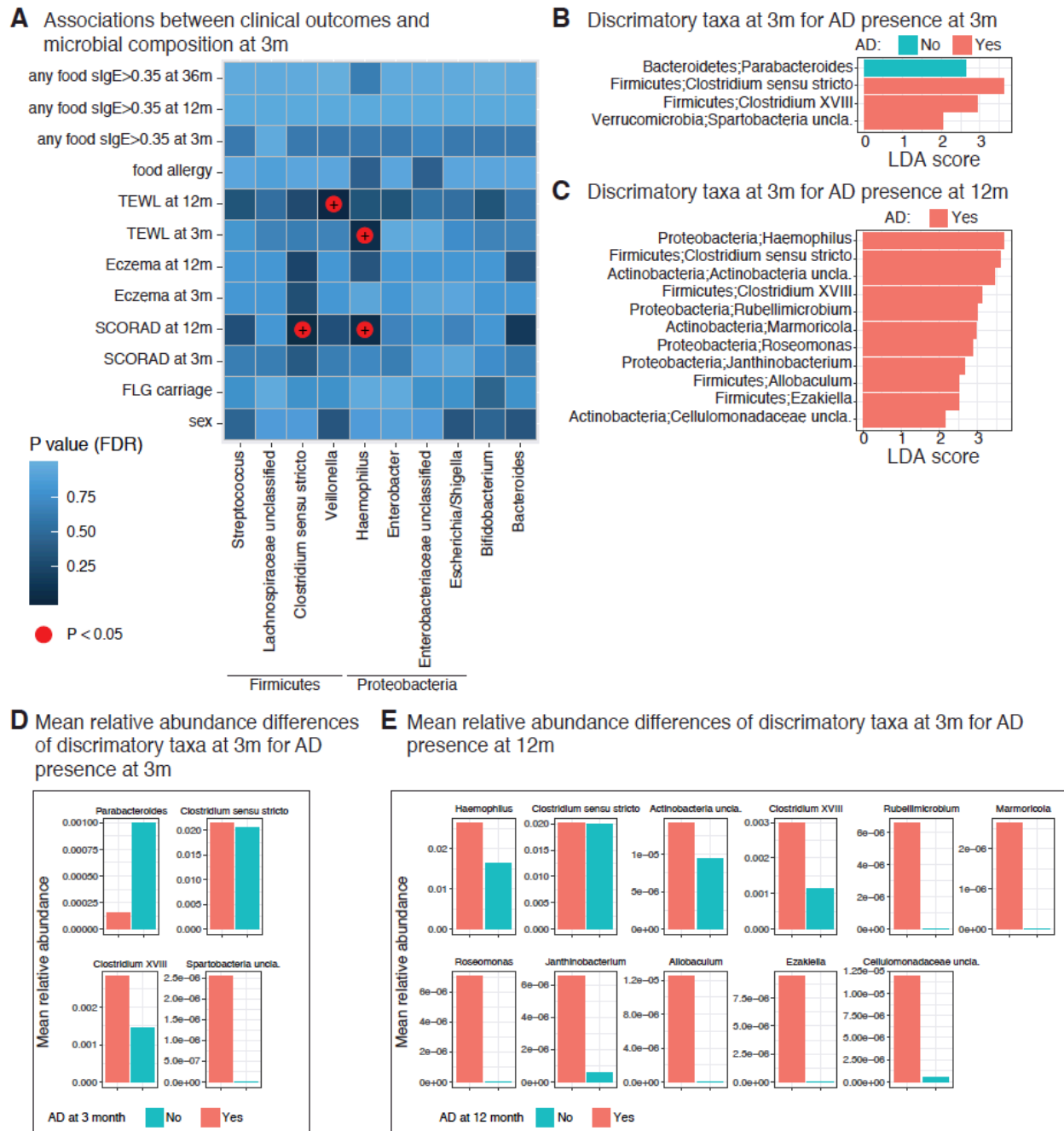


FIG E6. Association of clinical measurements with microbial compositions.

A, Heatmap showing associations between various clinical measurements and microbial composition. Darker colors indicate lower P values and red marks indicate associations with FDR-corrected P values lower than 0.05.

B, Discriminatory taxa from three-month gut microbiota in infants with or without eczema at three months of age (LDA score: Linear discriminant analysis score). LDA score is calculated by LEfSe(1).

C, Discriminatory taxa from three-month gut microbiota in infants with or without eczema at twelve months of age. LDA score is calculated by LEfSe.

D. Mean relative abundance differences of selected taxa from B (with or without eczema at three months).

E. Mean relative abundance differences of selected taxa from C (with or without eczema at twelve months).

FIG E7

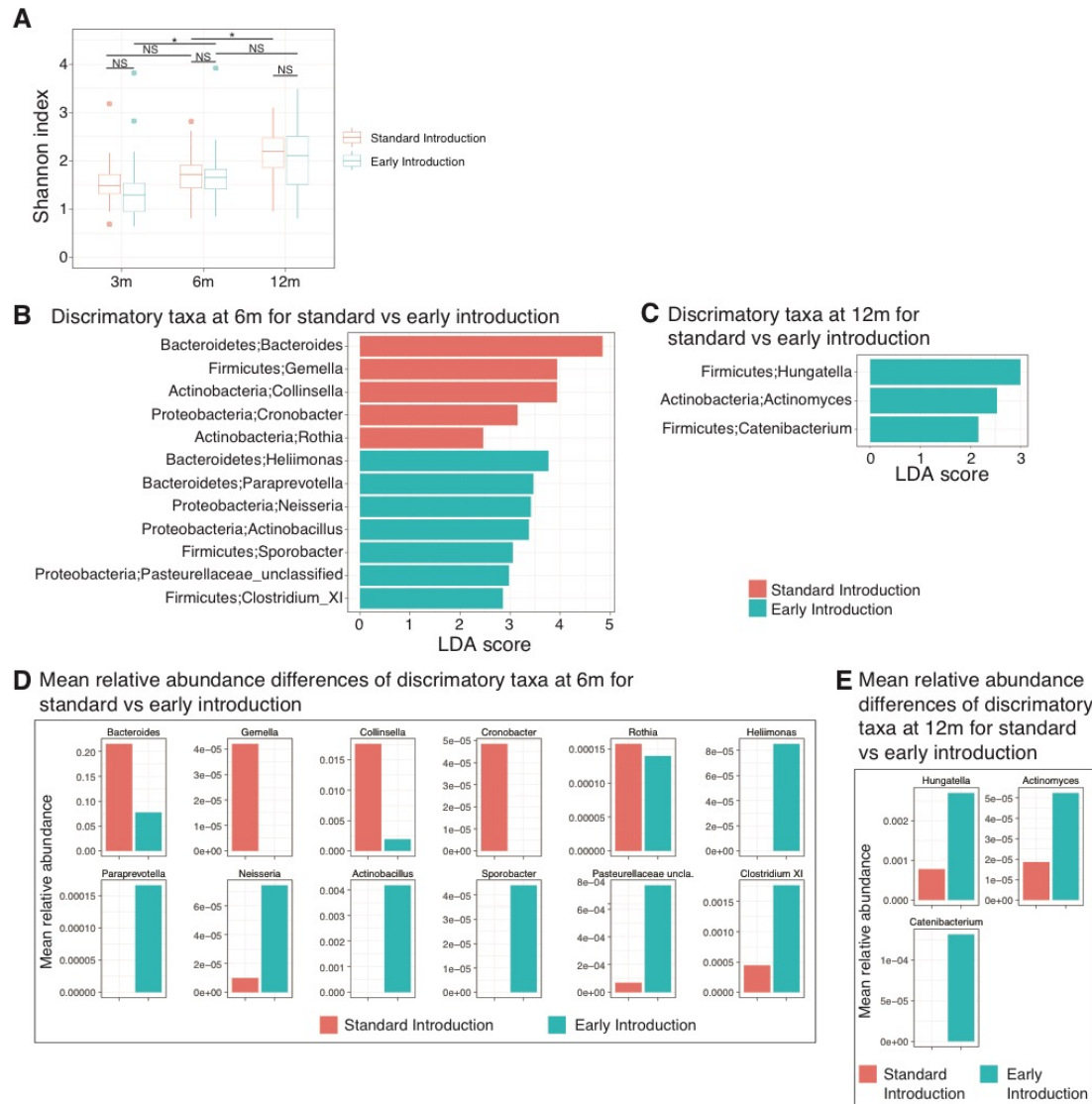


FIG E7. Differential microbiota dynamics by early food introduction in EAT cohort

A, Boxplot comparing Shannon diversity changes amongst participants' longitudinal samples according to randomized allocation to continued exclusive breastfeeding (standard introduction group) or the introduction of allergenic solids (early introduction group). (* $p < 0.05$; paired Wilcoxon rank-sum test)

B, Discriminatory taxa from six-month gut microbiota of infants in standard versus early introduction (LDA score: Linear discriminant analysis score). LDA score is calculated by LEfSe.

C, Discriminatory taxa from twelve-month gut microbiota of infants in standard versus early introduction (LDA score: Linear discriminant analysis score). LDA score is calculated by LEfSe.

D, Mean relative abundance differences of selected taxa from B (standard versus early introduction).

E, Mean relative abundance differences of selected taxa from C (standard versus early introduction).

FIG E8

A

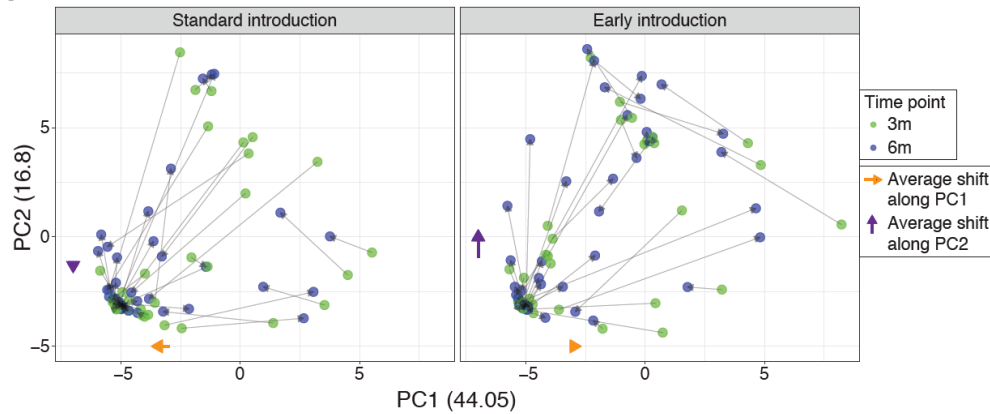


FIG E8. Differential microbiota dynamics by early food introduction in TEDDY cohort

A, PCoA scatter plot demonstrating longitudinal transition from 3 to 6 months of age of defined samples from the TEDDY cohort. From the TEDDY dataset samples from infants who were i) exclusively breastfed before solid food introduction and ii) longitudinal data (>2 sampling before 6 months of age) were selected. Grey arrows connect samples from the same individual. Yellow and purple arrows on the sides indicate the average shift of the microbiota in the PCoA axes.

Supplementary appendix

Supplementary methods

Chemical lysis was undertaken by adding 200µls of Guanidinium thiocyanate–EDTA–sarkosyl and 900µls of phosphate-buffered saline to the samples before they underwent bead-beating. Cell disruption was undertaken using Qiagen stainless steel beads and tungsten carbide beads on a Fastprep-24 Instrument (MP Biomedicals Europe, Illkirch, France) running at 6.5meters/second for 45 seconds. Two cycles of thermolysis followed by alternating incubation at 90°C and -20°C for 10 minutes each, before cell debris was pelleted by centrifugation at 13,000 g for 10 minutes. Supernatant was transferred to a fresh microfuge tube, where it was inverted with 140µl of 5 molar sodium chloride and 374µl 40% polyethylene glycol and precipitated for one hour at 4°C. DNA was pelleted by centrifugation at 13,000 g for 10 min and resuspended in 500µl of sterile distilled water. 300µl Phenol/chloroform (1:1) was added and samples were vortexed before centrifugation at 13,000 g for 5 minutes. The upper phase was then transferred to a fresh microfuge tube. Total DNA was then precipitated by the addition of an equal volume of isopropanol and 0.1 volume of 10 molar ammonium acetate and stored at -20°C for one hour. DNA was pelleted by centrifugation at 13,000 g for 10 minutes. Pelleted DNA was washed in 70% ethanol, dried and resuspended in 30µl of sterile distilled water (2, 3).

DNA extracts were amplified using universal bacterial primers targeting the 16S ribosomal RNA gene at hypervariable region V4 (515F-806R), which are tailed with sequences to incorporate Illumina (San Diego, CA) adapters and indexing barcodes (4). Sequencing was performed on the MiSeq instrument using version 2 chemistry and 250 cycles, stratifying the amplicon samples according to three, six and twelve month time points between each plate.

References

1. Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, et al. Metagenomic biomarker discovery and explanation. *Genome Biol.* 2011;12(6):R60.
2. Green GL, Brostoff J, Hudspith B, Michael M, Mylonaki M, Rayment N, et al. Molecular characterization of the bacteria adherent to human colorectal mucosa. *J Appl Microbiol.* 2006;100(3):460-9.
3. Rogers GB, Cuthbertson L, Hoffman LR, Wing PA, Pope C, Hooftman DA, et al. Reducing bias in bacterial community analysis of lower respiratory infections. *ISME J.* 2013;7(4):697-706.
4. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, et al. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc Natl Acad Sci U S A.* 2011;108 Suppl 1:4516-22.